MOLECULAR AND FUNCTIONAL ANALYSIS OF HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 ENV GENES

SARAH ASHELFORD

PhD

The University of Edinburgh

1996

ABSTRACT

The principal target of antibodies that can neutralise HIV-1 infection *in vitro* is the surface envelope protein, encoded by the envelope gene (*env*). The *env* gene evolves rapidly during the course of infection, producing a multitude of variants, many of which show different sensitivities to neutralisation by antibodies. The work described in this thesis, examines the hypothesis that the selection of antibody-resistant variants of the *env* gene is a strong force shaping the evolution of the *env* gene, a process that allows the virus to persist in the face of a strong immune response.

In order to examine the antibody-response to the *env* gene, 19 HIV-1 clones were constructed by replacing the *env* gene of the infectious molecular clone pHXB2-D, with homologous *env* sequences obtained from a single individual. The 19 *env* genes were obtained directly from patient peripheral blood mononuclear cell DNA using the polymerase chain reaction. The sequences of the *env* genes represent the predominant types found in the plasma at seroconversion and in five subsequent years of infection within this individual. No inactivating mutations were found following the partial sequencing of each *env* clone.

Following the introduction of the recombinant proviral clones into human T-cell cultures, only five of the 19 showed evidence of the production of high-replicating viruses. Viruses from these clones induced cell-to-cell fusion (or syncytia), and were capable of growth in a variety of human T-cell lines. With the remaining 14 clones, three showed evidence low level replication in human peripheral blood mononuclear cells, but failed to replicate in T-cell lines and did not induce the formation of syncytia. The remaining 11 clones but did not show evidence of productive infection.

In conclusion, despite being taken directly from an infected patient, only a small proportion of *env* genes were found to support productive infection *in vitro*.

ABBREVIATIONS

SI	syncytium-inducing
NSI	non-syncytium-inducing
MCS	multiple cloning sites
T-cell	T-lymphocyte
MAb	monoclonal antibody

CONTENTS

Chapter 1

INTRODUCTION

1.1 Impact of HIV on World Health

The human immunodeficiency viruses cause a chronic, debilitating infection which results in a profound loss of the immune system, culminating as the acquired immunodeficiency syndrome (or AIDS).

Although the first cases of AIDS were only reported as recently as 1981, HIV infection has caused an escalating pandemic. The World Health Organisation has estimated that in 1993, over 15 million people had been infected with HIV, and over three million had developed AIDS, most of whom have died. By the end of the century, the estimated number of cases of HIV infection, world-wide, is 30 to 40 million. In many large cities in the United States, AIDS is currently the commonest cause of death in young men. However, the situation is worse in the developing world, particularly in Sub-Saharan Africa and Southeast Asia, where the estimated number of cases is greater than 9 million and 2 million, respectively.

At present, there is no antiviral drug therapy capable of eliminating HIV-infection, nor is there a vaccine against HIV. Since its discovery in 1983, there has been an unprecedented effort of biomedical research, aimed at understanding the pathogenesis of HIV infection, the development of antiviral therapies, and vaccine research. Although HIV is one of the most intensively studied viruses, much remains to be understood of the pathogenic mechanisms of HIV infection from both a viral and immunological standpoint before effective treatment and vaccines can be developed.

1.2. The Natural History of HIV Infection

Although the course of HIV infection may vary between individual patients, a common pattern of disease development has been recognised, and may be divided into three stages: primary infection, asymptomatic infection, and symptomatic infection.

Primary Infection

The course of HIV infection begins with primary infection, which in the majority of cases (50 to 70%) is associated with an acute viral syndrome - a period of rapid virus replication, and the presentation of clinical illness which may include fever, rash, and in some cases hospitalisation may be required (Daar *et al*, 1991; Clark *et al*, 1991). However, many cases of primary infection will not produce any clinical symptoms.

During this period, the levels of infectious virus in the peripheral blood, can reach 10^3 cell infectious doses per 10^6 cells, with plasma-associated HIV RNA levels of greater than 10^7 molecules/ml (Piatak *et al*, 1993). Viral core antigen (p24) can be detected slightly after the detection of plasma viraemia and may reach levels of 10^2-10^3 pg/ml (Clark *et al*, 1991). There follows a rapid decline in plasma viraemia, coincident with the first detectable anti-p24 antibodies. Cytotoxic T-lymphocytes have been detected as the first antiviral immune response (Koup *et al*, 1994), with neutralising antibody responses occurring later, from between two weeks, to as long as 11 months after primary infection (Albert *et al*, 1990, Arendrup *et al*, 1992, Ariyosha *et al*, 1992 and Moore *et al*, 1994b).

Asymptomatic Infection

Following the down-regulation of virus production in the peripheral blood, the patient enters a period of clinical latency, or asymptomatic infection. During this period, viral infection appears to be confined to the lymphoid tissues and peripheral blood - where viral replication can be detected (Pantaleo *et al*, 1993; Michael *et al*, 1992; Zhang *et al*, 1991, Ho *et al*, 1989).

During the course of asymptomatic infection, the viral load increases, and the number

of CD4 T cells slowly declines (Levy, 1993).

The mean duration of the asymptomatic stage of infection is 10 years (Rutherford *et al*, 1990). However, in a very small number of patients, there is no period of clinical latency, and AIDS may develop within weeks of primary infection. At the other extreme, some patients, termed long term survivors, have a more protracted asymptomatic stage, in which levels of CD4+ T-lymphocytes may remain stable and within the normal range for fourteen years or longer (Haigwood *et al*, 1993; Cao *et al*, 1995).

Symptomatic Infection

After the asymptomatic phase of infection, the patient begins to show constitutional symptoms of viral infection - characterised by fever, night sweats, weight loss and in some cases, neurological disfunction. The hall-mark of AIDS is a dramatic reduction in the number of circulating CD4-positive T-lymphocytes, however, other immune dysfunctions can be detected (Miedema, 1992; Levy, 1993). The resultant profound immunodeficiency results in the appearance of opportunistic infections and the development of neoplasms. Death usually occurs within two years of developing AIDS.

During the period of symptomatic infection, viral replication can increase to 10^6 RNA copies/ml in the peripheral blood (Piatak *et al*, 1993, Zhang *et al*, 1991). At postmortem, high levels of virus can be detected in both lymphoid and non-lymphoid tissues (Donaldson *et al*, 1994a).

The Centers for Disease Control (CDC), has introduced, mainly for epidemiological purposes and clinical reporting, a four stage classification system for HIV-induced disease:

CDC I, acute infection;
CDC II, asymptomatic infection;
CDC III, persistent generalised lymphadenopathy;
CDC IV; other clinical symptoms of HIV-infection.

Groups I, II and III, are regarded as pre-AIDS conditions, which may or may not, progress to AIDS. Group IV, is subdivided into five groups, of which a subset are considered to constitute frank AIDS. In 1993, the CDC expanded its definition of AIDS to include HIV-infected persons with CD4 lymphocytes counts of less than 200 per ul.

1.3. The Virus

1.3.1. HIV Classification

The human immunodeficiency viruses belong to the **retrovirus** family, and to the **lentivirus** subfamily.

Retroviruses are characterised by the possession of a single-stranded RNA genome which is 'reverse-transcribed' into a DNA copy, through the virus-encoded enzyme reverse-transcriptase (Weiss *et al*, 1985b; Varmus, 1988). However, the possession of a reverse-transcriptase is not unique to retroviruses, the so-called 'para-retroviruses', such as the hepatitis B virus, also contain a reverse-transcriptase (- the replication of the double-stranded genome of Hepatitis B Virus occurs through an RNA intermediate which is reverse-transcribed). Retroviruses are distinguished additionally, by virion morphology and composition, and by the structure of the viral genome.

Historically, the retrovirus family has been subdivided into three taxonomic groups, based primarily on the disease manifestations caused on infection of the host, and on the *in vitro* consequences of infection. These groups are:

- the **oncovirus** subgroup, specific members of which can cause neoplastic disease in the infected host. This group includes Rous sarcoma virus, murine and avian leukaemia viruses, human and bovine T-cell leukaemia viruses;
- the **spumavirus** subgroup, which produce a 'foamy' cytopathic effect *in vitro* but are not clearly associated with disease *in vivo*. This group includes the human foamy virus;

- the **lentivirus** subgroup, of which, specific members may cause a chronic infection in which the disease state emerges in a *slow* fashion (ie. after many years of infection). This group includes the Equine Infectious Anaemia Virus (EIAV), Visna Virus of sheep, the closely related Caprine Arthritis Encephalitis Virus (CAEV) and the recently-discovered immunodeficiency viruses - which are found in many mammalian groups: human, simian, bovine, and feline.

There are two distinct types of the human immunodeficiency virus: type 1 (HIV-1) whose infection is found globally, and type 2 (HIV-2), whose infection appears to be confined to Western Africa. Both HIV subtypes can lead to AIDS, although there is evidence that the pathogenic course of HIV-2 is longer (Marlink *et al*, 1994).

Despite their name, a number of lentiviruses can cause a rapid, acute disease in the host, especially under experimental conditions infecting a new host species. For example, infection of pigtailed macaques with an highly pathogenic strain of SIV from sooty mangabeys ($SIV_{SMM\text{-}PBj14}$) can cause death within eight days of inoculation (Fultz *et al*, 1989). Primary infection with the human immunodeficiency viruses frequently results in an acute viral syndrome - which is then followed by a more characteristic, slow progressive disease.

It has been argued that for comparative purposes, the retrovirus family should be classified according to the complexity of the retroviral genome. In this way common patterns of viral life-cycle and modes of infection and pathogenesis may be recognised. Cullen, 1991, has suggested that the retrovirus family be classified into two groups: **simple** and **complex** retroviruses. Simple retroviruses, such as the murine leukaemia virus, contain three genes: *gag*, *pol* and *env*, common to all retroviruses. Complex retroviruses contain additional regulatory genes, allowing a more complex pattern of gene expression. All lentiviruses and the human T-cell leukaemia viruses would fall into this latter group.

The morphology of the virus, as viewed under the electron microscope, has also been used to subclassify the retrovirus family. In this system, four subgroups have been identified: **types -A, -B, -C** and **-D**. The members of each subgroup have a

characteristic extra-cellular (virion) and intra-cellular form. This classification system is consistent with the genetic classification and has allowed the relatively rapid identification of unknown viral isolates (Weiss *et al*, 1985b).

1.3.2. The Virion

HIV is a small, enveloped virus, the characteristic features of which are depicted in figure 1.1.

The virion contains an inner core, or capsid (CA), whose cone-shape is characteristic of the lentivirus subfamily (Gelderblom, 1991). The capsid contains the viral nucleic acid, which is present as two copies of positive-sense RNA, and, in addition, contains the viral proteins necessary for the early replication steps, such as the reverse-transcriptase (RT) and integrase (IN) enzymes. The capsid is surrounded by a membrane-associated matrix protein (MA), which is thought to stabilise both the capsid and the viral envelope. The virion envelope is a lipid-bilayer derived from the host cell during the process of viral budding, and contains the virus-encoded envelope surface (SU) and transmembrane (TM) proteins, and also contains host-cell-derived proteins, in particular, there is an abundance of the immune-system molecules class I and II HLA and β2-microglobulin (Arthur *et al*, 1993).

Figure 1.1 A Schematic Representation of the HIV-1 Virion

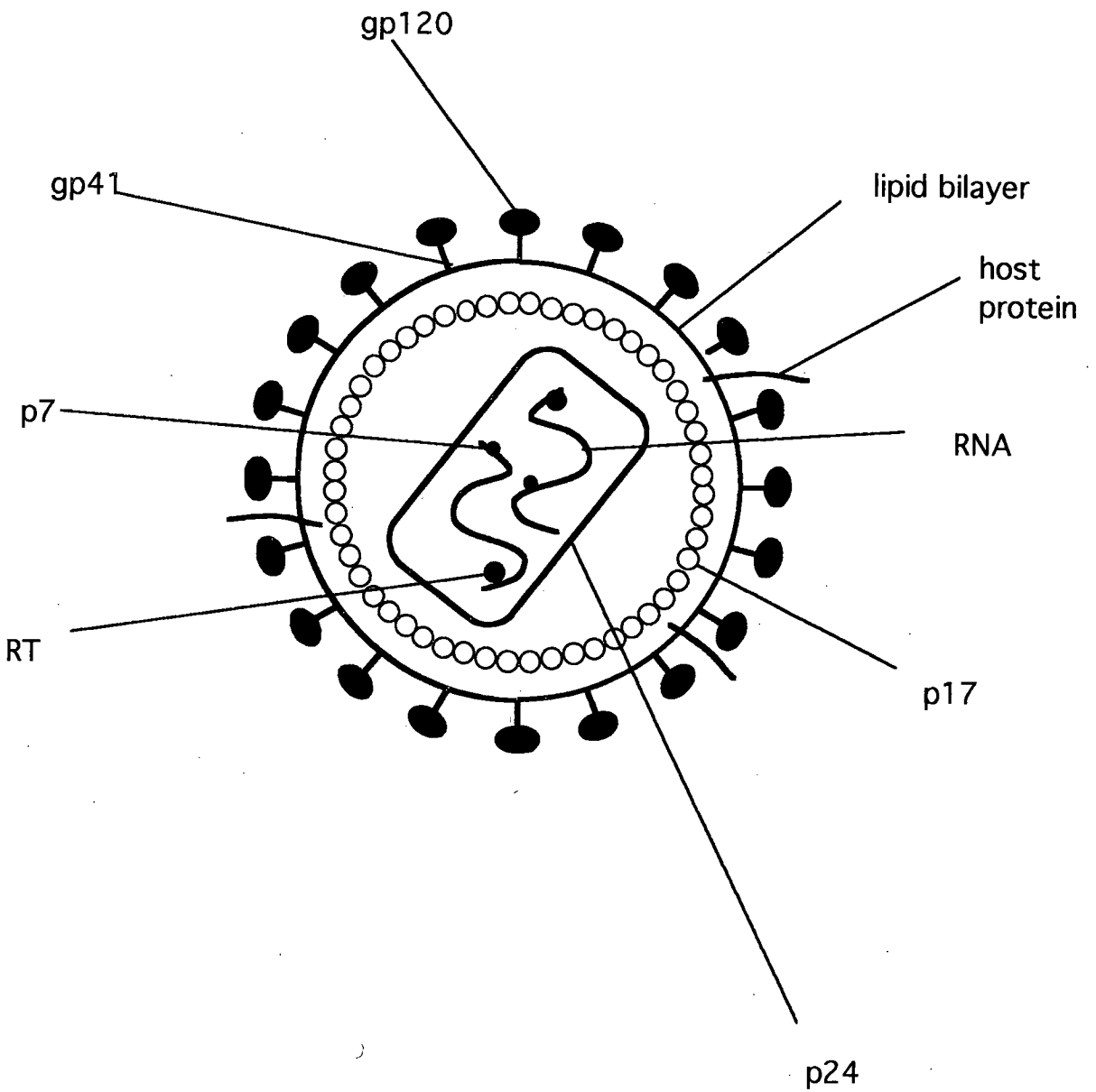
gp41 and gp120 - virus envelope proteins; p7, p17, p24 - viral core proteins

RT - virus reverse transcriptase enzyme; RNA - positive-sense, diploid viral RNA genome

lipid bilayer - virion envelope derived from the host cell

host proteins - include MHC class I and II molecules and β-microglobulin

opposite page 8


gp120
gp41
lipid bilayer
host
protein
p7
RNA
RT
p17
p24

1.3.3. The HIV Life-Cycle

The life-cycle of HIV and other lentiviruses, is similar to that of the more simple retroviruses (Varmus, 1988). The main differences reside in the regulation of viral gene expression; lentiviruses have additional genes which are potent regulators of gene expression. In HIV-1 these are the *tat* and *rev* genes (Felber and Pavlakis, 1993).

The first step occurs when the virion encounters and binds to a susceptible host cell - the CD4+ T lymphocyte, or the monocyte/macrophage - through a high-affinity receptor on the host cell, which for HIV-1 is the CD4 molecule (Dalgleish *et al*, 1984; Klatzman *et al*, 1984b; Maddon *et al*, 1984). Two second receptors have been identified recently: CC-CKR-5 and fusin (Dragic *et al*, 1996; Deng *et al*, 1996; Feng *et al*, 1996). CC-CKR-5 is required for entry of certain strains of HIV-1 (those with a macrophage and T-cell tropism). Fusin is required for entry into the cell of T-cell line-tropic strains of HIV-1. This is discussed further in sections 1.3.10 and 1.4, below.

Following binding, the virion is internalised by a process of direct-fusion of virus and cell membranes, or by receptor-mediated endocytosis and subsequent fusion of membranes intra-cellularly (Maddon *et al*, 1984; McClure *et al*, 1988) .

The virion is then uncoated by the complete or partial loss the envelope proteins and virion membrane. The virus particle remains in the host-cell cytoplasm as a nucleoprotein complex, where the first stages of reverse-transcription occur. The diploid, single-stranded RNA genome is copied into a (double-stranded) DNA copy, through the complex process of reverse-transcription.

In the activated, proliferating T-cell, the nucleoprotein (pre-integration) complex is then translocated into the host-cell nucleus, and the double-stranded genome maybe inserted into the host-cell genome, where it resides as a **provirus**.

The provirus is transcribed by the host-cell RNA polymerase II and associated transcription factors. The early gene products are the virus regulatory proteins (Tat,

Rev and Nef) which in turn, promote the synthesis of structural, enzymatic and maturation proteins of the virion (Gag, Pol, Vif, Vpu, Vpr, and Env). New viral particles are assembled below the host cell membrane, and released by budding from the cell.

Most retroviruses depend upon cell proliferation for replication. At mitosis the nuclear envelope breaks down allowing the pre-integration complex to interact with the host-cell genome (Lewis and Emerman, 1994). However, lentiviruses can infect non-replicating cells, such as macrophages and resting T-cells. HIV-1 contains nuclear localisation signals on the viral matrix (p17) and Vpu proteins, which directs the pre-integration complex into the nucleus (Bukrinsky *et al*, 1993; Heinzinger *et al*, 1994; Zack *et al*, 1990). In terminally-differentiated macrophages, and cells arrested at G_1/S or G_2, reverse-transcription, integration and viral replication may then occur (Weinberg *et al*, 1991).

However, following infection of resting T-lymphocytes - those at G_0 of the cell cycle - complete reverse-transcription does not occur, and extra-chromosomal partial reverse-transcripts can be found (Zack *et al*, 1990; Stevenson *et al*, 1990a). These may direct the expression of *gag* and *env* genes (Stevenson *et al*, 1990b), although this has not been confirmed. Following T-cell activation, integration, and viral gene expression can occur.

1.3.4. The HIV-1 Genome

The unusual life-cycle of the retrovirus involves two forms of the viral genome: a diploid, single-stranded RNA form, which is present in the extra-cellular virion, and a double-stranded DNA, or proviral form - which describes the viral genome when it is integrated into the host-cell genome. The proviral form also exists as intra-cellular, unintegrated, circular or linear forms, of which the latter is the normal pre-integration intermediate (Weiss *et al*, 1985b).

The process of reverse-transcription results in the partial duplication of *cis*-acting control elements found at the ends of the RNA form necessary for viral replication (elements 'R' and 'U5' at the 5' end, and 'U3' and 'R' at the 3' end). This produces

(elements 'R' and 'U5' at the 5' end, and 'U3' and 'R' at the 3' end). This produces two identical structures (or long-terminal repeats) at the ends of the provirus, containing both U3, R and U5 (figure 1.2).

Genetic Elements Other Than Coding Regions

The long terminal repeats (LTRs) contain the essential *cis*-acting elements defining the start-site of RNA-transcription and the polyadenylation site, common to all mRNA species, and elements controlling the level of RNA expression. These elements include the binding-sites of ubiquitous cellular transcription factors Sp1 and TFIID - the Sp1 binding site and the TATA box (which are common to all retroviruses) (Garcia *et al*, 1987; Jones *et al*, 1986). The LTRs also contain binding sites for the inducible host transcription factor NF-κB, whose activity is induced following stimulation of the cell (by mitogens and antigens; Nabel and Baltimore, 1987; Gruters *et al*, 1991). The LTR also contains the TAR element (TransActivator Response region), through which the viral trans-activator protein, Tat, mediates its effect.

Other recognised genetic elements include a multitude of RNA splice -donor and -acceptor sites, of which there are several sets for the smaller regulatory genes (Schwartz *et al*, 1990a, 1990b). There is the Rev-Response Element, or RRE, lying at the 3' end of the genome, which is a regulatory element associated with the viral Rev protein.

Coding Sequences

The HIV-1 genome encodes nine genes, whereby 'gene' refers to the region encoding a primary protein product. Six of the nine genes are referred to as 'essential genes', each of which is required for growth of the virus *in vitro*. These are *gag*, *pol* and *env* genes, which are common to all retroviruses, and the three regulatory genes *vif*, *tat* and *rev*. There are, in addition, three 'non-essential genes'; *vpr*, *vpu* and *nef*, which are not required for establishing viral cultures, although some may be required for growth in specific cell culture systems (section 1.4.6, below).

Figure 1.2 The HIV-1 Genome

shaded boxes represent open reading-frames

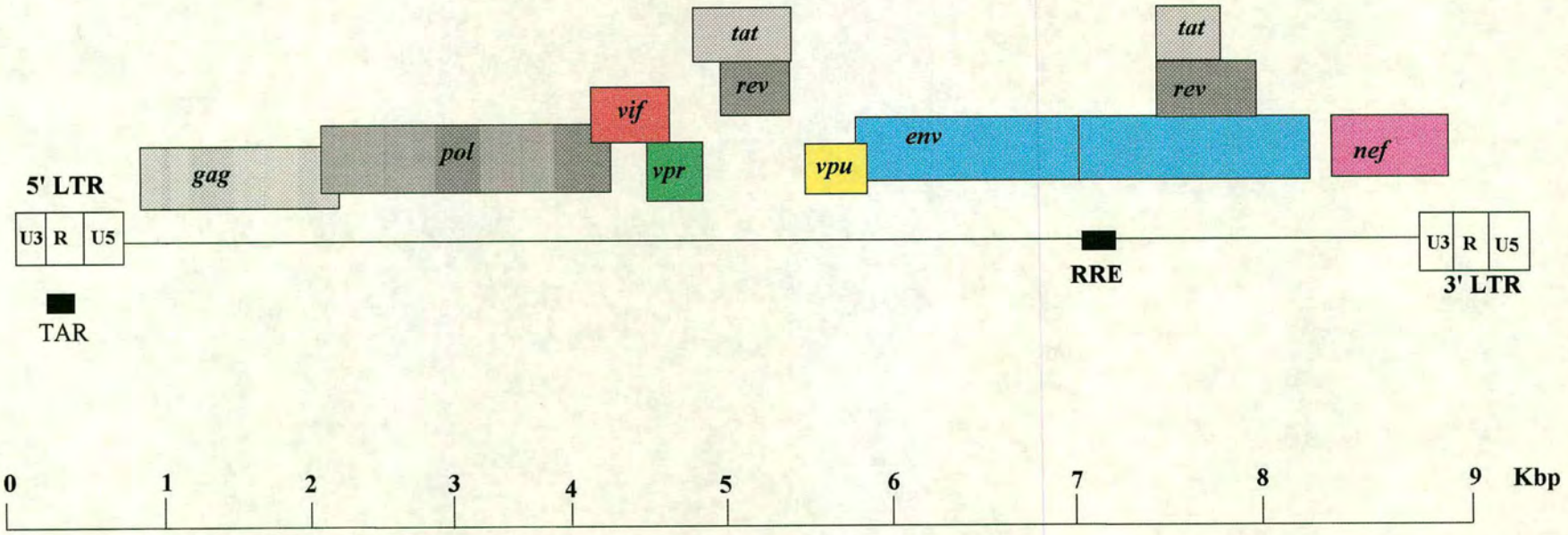
TAR transactivator response region

RRE Rev-responsive region

LTR long terminal repeat

opposite page 12

The HIV-1 Genome



1.3.5. Non-Human Primate Immunodeficiency Virus Genome Organisation

The experimental infection of various non-human primate (simian) species with SIV or HIV, arguably produce the most relevant animal models in which to investigate the pathogenesis of HIV, or SIV infection, and with which to carry out vaccine research. Based upon genetic sequence analysis, used to determine the relatedness of the primate lentiviruses, four discrete groups are now recognised (Myers *et al*, 1992):

- HIV-1/SIVcpz (wild caught chimpanzee) group;
- SIVsmm (captive and feral sooty mangabeys)/HIV-2/SIVmac (captive macaques);
- SIVagm (wild caught African green monkeys);
- SIVmnd (wild caught mandrills).

Together, the four groups contain seven 'auxiliary' genes, which are not found in other retroviruses. These are: *tat*, *rev*, *nef*, *vif*, *vpr*, *vpu* and *vpx*.

- four out of the seven are common to all four primate lentivirus groups: *tat*, *rev*, *nef* and *vif*.
- *vpu* is unique to the (HIV-1/SIVcpz) group.
- *vpx* is found in the SIVagm and (SIVsmm/HIV-2/SIVmac) groups.
- *vpr* is found in the (HIV-1/SIVcpz) group, the (SIVsmm/HIV-1/SIVmac) group, and the SIVmnd group.

Nucleotide similarity suggests that *vpx* is a duplicated *vpr* gene. Thus the absence of *vpr* in the SIVagm group may be compensated by *vpx*, and the absence of *vpx* in the (HIV-1/SIVcpz) group may be compensated by *vpr*. *vpu* stands out as unique to the (HIV-1/SIVcpz) group (Myers *et al*, 1992).

1.3.6. The Pattern of HIV-1 Gene Expression

HIV-1 produces a complex set of mRNA molecules from the alternative splicing of a full-length precursor mRNA, initiated from a common position in the 5' LTR (figure 1.3).

The transcripts may be classified into two groups, according to the regulation of their expression, and the stage of the viral life-cycle at which they are produced:
i) unspliced and partially-spliced, RRE-containing; these are Rev-dependent for their expression. These transcripts are produced late in the viral life-cycle, and include the messages for the *gag*, *pol*, *vpu*, *vpr*, *vif* and *env* genes.
ii) small, multiply-spliced mRNAs, that lack the RRE and are Rev-independent. These transcripts are produced early in the viral life-cycle and include the messages for the regulatory genes *tat*, *rev*, and *nef*.

The pattern of gene expression is also regulated at the protein level, with many transcripts encoding for more than one protein, thus maximising the coding potential of the viral genome. Differential gene expression at the protein level may, for example, be the result of ribosome frameshifting (*gag* and *pol* expression; Wilson *et al*, 1988), or the use of alternative translation start-sites (*vpu* and *env*; Schwartz *et al*, 1990b). Post-translational proteolytic cleavage also results in the production of multiple proteins from single polypeptide precursors, as is seen with the Gag, Pol and Env proteins.

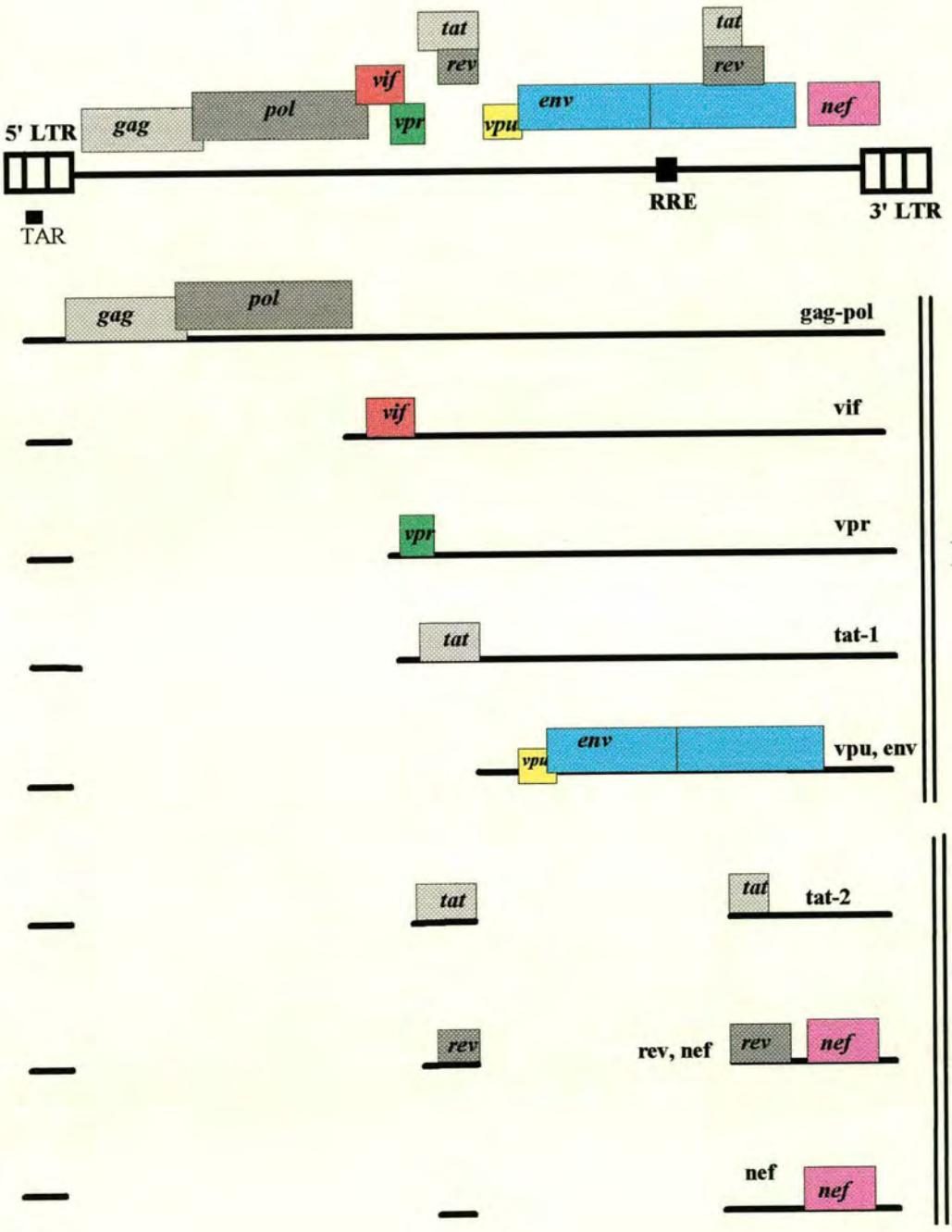
Figure 1.3 Principal mRNAs Produced from the HIV-1 Genome

This figure is adapted from Schwartz *et al*, 1990a, 1990b.

Shaded boxes represent open reading-frames
RRE+ mRNAs that contain the RRE, and are Rev dependent for their expression
RRE- mRNAs that have the RRE spliced out, these are Rev-independent for their expression.

opposite page 15

HIV-1 mRNAs


tat
rev
vif
vpr
tat
rev
5' LTR
gag
pol
vpu
env
nef
RRE
3' LTR
TAR
gag
pol
gag-pol
vif
vif
vpr
vpr
RRE+
tat
tat-1
env
vpu
vpu, env
tat
tat
tat-2
rev
rev, nef
rev
nef
RRE-
nef
nef

1.3.7. The Function of the Genes of HIV-1

- gag

In common with other retroviruses, the Gag and Pol proteins are translated from an unspliced *gag-pol* transcript (figure 1.3), with Pol translation resulting from ribosome-frameshifting (Wilson *et al*, 1988). The *gag* gene encodes a 53kD Gag precursor polypeptide ($p55^{gag}$) which is cleaved by the virus-encoded protease to produce the p24 capsid (CA) protein, the p17 matrix (MA) protein, and p9 and p7 Gag proteins of ill-defined function (figure 1.1).

The matrix protein, p17, contains a nuclear localisation signal, which directs the pre-integration complex to the nucleus. This allows the establishment of HIV-infection in macrophages, which are terminally-differentiated cells, and resting T-cells (von Schwedler *et al*, 1994; Bukrinsky *et al*, 1993).

- pol

The *pol* gene encodes a Pol precursor protein which is cleaved sequentially to yield the reverse-transcriptase (RT), protease (PR) and integrase (IN) enzymes. The reverse-transcriptase contains a DNA-dependent DNA polymerase, an RNA-dependent DNA polymerase and an RNaseH activity, each of which is utilised during reverse transcription of the RNA genome into a double-stranded DNA version. RT exists as a heterodimer containing a p66 and p55 subunit from the same coding region. The integrase is responsible for the integration of the viral DNA into the host genome to form the provirus, and the protease is responsible for the post-translational modification (maturation) of the viral proteins.

- tat

The Tat protein is an 86 amino acid nuclear protein, that acts as a powerful trans-activator of HIV gene expression, increasing the efficiency of transcription by over 1000-fold (Marciniak *et al*, 1990). Tat is required for replication of HIV *in vitro*, with mutations in *tat* destroying infectivity (Fisher *et al*, 1986). The mechanism of the action of Tat is unresolved, but appears to be primarily by preventing premature

termination, through binding to the RNA transcript itself (Laspia *et al*, 1989). *Trans*-activation by Tat requires a *cis*-acting target sequence known as the trans-activation response element, or TAR element, which is situated just down stream of the transcription initiation site, and is therefore common to all HIV transcripts (figures 1.2 and 1.3). The TAR element forms a stem-loop structure with which Tat, in association with cellular factors, interacts directly to exert trans-activation.

- rev

The Rev protein controls the pattern of HIV mRNA expression, promoting the transition from 'early' regulatory gene expression to 'late' structural gene expression (Malim *et al*, 1988). Rev is an early gene product, translated from multiply-spliced mRNA (figure 1.3), and is essential for viral replication *in vitro*. Experiments using *rev*-defective proviruses show that Rev is required for the detection of unspliced mRNA and singly spliced RNA. Analysis of the localisation of HIV-1 RNA reveals that unspliced and singly-spliced mRNA are retained in the nucleus (Chang and Sharp, 1989). Thus it appears that Rev is required for transport of these late mRNAs into the cytoplasm.

The mechanism by which Rev acts is not clear. However, like Tat, Rev exerts its effect on the mRNA itself, through a *cis*-acting target sequence, the Rev-responsive-element, or RRE, which is situated within the *env*-coding region (Heaphy *et al*, 1990, Malim *et al*, 1990, 1991). The RRE is therefore present in the late mRNAs, but is spliced out of the early transcripts (figure 1.2). The late transcripts, containing the RRE, require the presence of Rev for their expression, ie. are RRE/Rev dependent. The early transcripts do not contain the RRE, and are therefore Rev/RRE independent.

- tev

The *tev* transcript has been identified as a minor HIV transcript from analysis of *in vitro* cultures of certain HIV-1 strains (for example, the HIV-1 strain HIV-1_{IIIB} Shaw *et al*, 1984). However, sequence analysis has shown that *tev* is not present in all HIV strains. Tev protein has both Tat and Rev activities and can replace both these essential genes (Benko *et al*, 1990).

- ***vif***

The Vif protein is a late-gene product, produced from singly-spliced mRNA (figure 1.3). The *vif* gene is present in all four groups of primate lentiviruses, implying the importance of this gene. *vif* mutants of HIV-1, have dramatically-reduced replicative capacity *in vitro*, the magnitude of which, is dependent upon the cell system used (Gibbs *et al*, 1994). The mechanism of action of the Vif protein is not known, but it appears necessary for cell-free viral infection, but not for direct cell-to-cell transmission. Vif appears to play a role in the formation of virus particles, however, Vif itself is only found at trace amounts in the virion. *Vif*-defective viruses produce virions, but these have an improperly packed nucleoprotein core (Hogland *et al*, 1994) and fail to synthesise proviral DNA (von Schwedler *et al*, 1993).

- ***vpu***

Vpu is a 16kD protein, expressed from a bi-cistronic mRNA along with the *env* coding region (Strebel *et al*, 1988, Schwartz *et al*, 1990b), which implies a role for the co-ordinate expression of the *env* and *vpu* genes. Despite this, Vpu is dispensable for viral replication *in vitro*. The main action of *vpu* in T-lymphocyte culture appears to be to promote the export of viral particles from the cell, but Vpu does not have a profound effect on the replication rate in this cell type (Klimkait *et al*, 1990). However, it appears that Vpu may be essential for virus growth in cultures of monocyte-derived macrophages, where *vpu*-defective viruses show a block to replication, or severely impaired growth, but may be rescued by co-cultivation with T-lymphocytes (Westervelt *et al*, 1992b, Kawamura *et al*, 1994). The role of *vpu* on growth in monocyte-derived macrophage cultures appears to be complementary to role of *vpr*, such that one of these genes need be functional for replication in this cell type.

Two apparently unlinked roles have been described for the Vpu protein; firstly, a CD4-independent role, in which Vpu is involved in promoting the release of virus particles from the plasma membrane (Terwilliger *et al*, 1989, Klimkait *et al*, 1990), and secondly, a role in the intra-cellular degradation of CD4, with a consequent reduction in the delivery of CD4 to the plasma membrane (Geraghty *et al*, 1993). The down-regulation of the receptor molecule is a common feature of retroviral infection,

where its role may be to prevent super-infection by additional viruses.

In the infected cell the HIV-1 envelope precursor protein, gp160, binds to the CD4 molecule in the endoplasmic reticulum (ER), forming a gp160-CD4 complex, which results in a reduced delivery of CD4 to the cell surface. Vpu appears to induce the degradation of CD4 within the gp160-CD4 complex, releasing gp160, which is subsequently processed into gp120 and gp41 and delivered to the cell surface (Bour *et al*, 1995; Lenburg and Landau, 1993). The biological significance for HIV-1 infection *in vivo* of the down-regulation of CD4 and the release and processing of gp160, however, remain to be elucidated. It is of interest that *vpu* is absent from the three other primate immunodeficiency virus groups (section 1.3.3), whose envelope proteins have been shown to have a lower affinity for CD4 than the HIV-1 envelope protein (Moore, 1990). This suggests that HIV-1 may have acquired the Vpu protein to prevent the formation of the CD4-gp160 complex, thereby releasing a block to gp120-gp41 production.

- ***vpr***

Vpr is a 12-15kD protein synthesised from a singly-spliced mRNA, late in infection. The Vpr protein is incorporated into the virion, where it is associated with the Gag precursor protein ($p55^{gag}$). Vpr contains a nuclear localisation signal to direct the pre-integration complex into the nucleus. The connection to the nuclear import pathway is necessary for viral infection of non-dividing cells: macrophages and resting T-cells (Heinzinger *et al*, 1994).

Vpr is a non-essential gene, in that laboratory-adapted viruses defective in *vpr* are able to replicate in T-lymphocyte cultures, with equal efficiency to vpr^+ viruses. Some researchers have found that *vpr*-defective proviruses are unable to grow in cultures of the monocyte-derived macrophage lineage (Westervelt *et al*, 1992b). Others, however, have found that *vpr*-defective proviruses show only a small decrease in the replication level in this cell type (Kawamura *et al*, 1994).

The requirement of *vpr* in the replication of HIV-1 in monocyte-derived macrophages appears to be complemented by the *vpu* gene, such that the presence of a functional

copy of either of these genes allows the viral replication in monocyte-derived macrophages culture (Westervelt *et al*, 1992b). The HIV-2 *vpr* gene has similarly been shown to be required for replication in monocyte-derived macrophage cultures, but dispensable for growth in T-lymphocytes (Hattori *et al*, 1990).

- ***nef***

The *nef* gene is found in all primate lentiviruses. Nef is dispensable for viral replication *in vitro* (Terwilliger *et al*, 1986), but appears essential for pathogenesis *in vivo* following the experimental infection of Rhesus macaques with SIVmac (Kestler *et al*, 1991).

The *in vivo* role of *nef* was demonstrated by experiments of Kestler *et al*, 1991, who showed that infection of Rhesus macaques with a *nef*-deleted clone of SIVmac results in low titre virus infection, and a stable CD4 count. In contrast, infections of macaques with the isogenic *nef*-intact SIVmac virus results in the rapid progression to a severe immunodeficient state, similar to the human AIDS. Furthermore, a strong selective pressure for *nef* function *in vivo* was shown by the reversion to a fully-encoding *nef* gene, of the previously *nef*-mutant viruses within two weeks of infection of the inoculated animals. These animals exhibited high levels of viraemia and subsequently developed AIDS.

The *in vitro* function of the Nef protein remains ill-defined; it gained its name as a negative factor for viral replication. Others, however, have found *nef*-defective viruses to replicate more rapidly, or at least equally-well as *nef*-intact viruses (Kin *et al*, 1989).

The *nef* gene from many isolates appears to down-regulate the surface levels of the CD4-receptor molecule (Mariani and Skowronski, 1993). Aiken *et al*, 1994, have shown that the Nef protein promotes the endocytosis of CD4, where it is subsequently degraded in lysosomes. The down-regulation of CD4 by Nef occurs at early stage, prior to the envelope protein-mediated receptor interference (described under the section on *vpu*, above).

Nef also appears to enhance virus infection and replication, a function independent of CD4 down-regulation (Miller *et al*, 1994; Spina *et al*, 1994); this may arise by enhancing the reverse-transcription process (Schwartz *et al*, 1995).

- env

The *env* gene of HIV-1 encodes the viral envelope (ENV) proteins. The viral envelope proteins play a pivotal role in the viral life cycle: the initial binding of the virion to the host cell, and subsequent internalisation of the virion. The envelope also plays a role in cell-to-cell transmission of the virus, and is the major target of antibodies capable of neutralising viral infectivity.

The *env* gene is initially translated as a glycoprotein precursor of molecular weight 160 kdaltons, designated gp160. gp160 is cleaved intra-cellularly to produce the surface (SU) glycoprotein gp120, and the transmembrane (TM) protein gp41, a step essential for viral infectivity (McCune *et al*, 1988, Willey *et al*, 1988). Both the precursor and mature cleavage products are highly glycosylated, with almost half the native molecular weight of gp160 and gp120 attributable to asparagine- (N-) linked carbohydrate (Putney, *et al*, 1986, Leonard *et al*, 1990).

After their cleavage, gp120 and gp41 are exported to the surface of the cell where they associate via non-covalent interactions. On the surface of the virion, the gp120-gp41 complexes are associated as heterodimers, forming oligomeric spike structures that are visible under the electron microscope (Gelderblom, 1991). The strength of the gp120-gp41 association varies between different virus isolates. For cell-line adapted strains, such as HIV-1IIIB (Shaw *et al*, 1984), the association is weaker than for primary isolates. gp120 can be lost from the surface of the virion, a property that correlates with loss of infectivity (McKeating *et al*, 1991).

1.3.8. The Structure of the HIV-1 Envelope Surface Protein gp120

The surface envelope protein, gp120, is folded into a complex three-dimensional structure, held together by disulphide bonds, whose number are invariant between widely divergent isolates (Myers *et al*, 1991). The assignment of the intra-chain disulphide bonds was carried out by Leonard *et al*, 1990, using a recombinant form of gp120, produced by expression in Chinese hamster ovary cells. This study has produced a definitive two-dimensional structure for gp120, which is depicted in figure 1.4.

The disulphide bonds delineate several domains of gp120, which are necessary for the correct folding of the molecule into a structure capable of binding the cellular receptor molecule CD4. The domains delineated by the disulphide bonds coincide with those defined as conserved or variable, according to sequence comparisons of divergent isolates of HIV-1 (Starcich *et al*, 1986, Modrow *et al*, 1987). Sequence comparisons have defined five **hypervariable** regions in gp120, designated V1 to V5 (figures 1.4 and 1.6). Interspersed between the hypervariable regions, are four conserved regions, designated C1 to C4.

The V1 and V2 hypervariable regions form a complex double-loop disulphide-bonded structure. The V3 and V4 hypervariable regions exist as single loop structures, formed by a single disulphide bond at the base of each. Computer-assisted modelling predicts that each of the hypervariable domains will be exposed at the surface of the molecule and potentially antigenic (Modrow *et al*, 1987). The V2 and V3 loops do indeed form important targets for antibodies found during the course of natural infection (section 1.6, below).

Figure 1.4

(A) The HIV-1 Surface Envelope Protein gp120

This figure is taken from Leonard *et al*, 1990, and shows the position of the cysteine bonds, and sites of N-linked glycosylation determined for CHO-cell produced recombinant gp120, from the HXB2 clone.
Roman numerals (I to V) indicate cysteine-bonded loops.
Boxed regions indicate hypervariable regions (V1 to V5), defined by Modrow *et al*, 1987. N-linked glycosylation sites are shown.

(B) The HIV-1 Transmembrane Protein gp41

This figure is taken from Ratner, 1992 and shows the primary structure of the mature gp41 predicted for the HXB2 clone.
The region important for fusion (the fusion peptide), and the location of potential N-linked glycosylation are shown.

opposite page 23

The C4 region has been shown to contain residues important for binding the receptor molecule, CD4 (Lasky *et al*, 1987). However, the testing of panels of gp120 mutants for their ability to bind soluble, recombinant CD4 has shown that other regions of gp120 are important for binding CD4, especially sites C2 and C3 (Olshevsky *et al*, 1990).

The three-dimensional structure of gp120 has not yet been elucidated, due in part to the high glycosylation of this molecule, which has precluded the formation of a crystal form of gp120. In consequence, an immunochemical approach has been undertaken by many researchers to determine in more detail the structure of gp120 (Sattentau *et al*, 1989, Moore *et al*, 1993a, 1993b, 1994a).

This approach has revealed that the conserved regions of gp120 are inaccessible to anti-gp120 monoclonal antibodies (MAbs) on the native protein. In contrast, the variable regions of the envelope are accessible to MAbs and are therefore exposed on the protein surface, although MAbs were found to be frequently masked by carbohydrate moieties (Moore *et al*, 1994a). In addition, the C1 and C5 portions of gp120, which are located at the amino terminal and carboxyl-terminal of gp120, respectively, were found to be exposed on a monomeric form of gp120. However, when complexes of gp120 and gp41 were probed with MAbs, these regions were found to be inaccessible, implying that these regions are hidden by contact with gp41 (Moore *et al*, 1994a).

The examination of panels of recombinant gp120 proteins altered by site-directed mutagenesis, has been a means of determining further the interactions of the different domains of gp120. Amino acid changes in the C1 domain of gp120 affect the binding of MAbs mapping to the C2 domain, implying a functional interaction between C1 and C2. Similarly, mutations in C4 enhance the access of MAbs to the V1/V2 structure. Mutations in C4 also affect the binding of MAbs to the V3 loop, suggesting that there is an interaction between C4 and both the V3 loop and the V1/V2 structure (Moore *et al*, 1993a, 1993b). There is also a structural relationship between the V3 loop and residues in C2 and C4, which contribute to the CD4-binding site (Wyatt *et al*, 1992).

1.3.9. The Structure of the HIV-1 Envelope Transmembrane Protein gp41

gp41 consists of three domains: a hydrophobic external domain at the N-terminal, a membrane spanning domain, and an intracellular domain which consists of an extensive cytoplasmic tail (figure 1.4B).

The external domain of gp41 interacts with gp120 via non-covalent bonds. Genetic data suggest that there are two or more contact points between gp120 and gp41 (Gallaher, 1987; Gallaher *et al*, 1989, Helseth *et al*, 1990b, Ivey-Hoyle *et al*, 1991). Computer modelling predicts that one of the contact-points in the surface proteins (SU) of divergent lentiviruses and oncoviruses forms a structurally-conserved pocket (Schulz *et al*, 1992). This pocket could accommodate a knob-like protrusion formed by the transmembrane (TM) immunodominant region. Thus the SU-TM linkage is predicted to be a 'knob and socket' structure.

The external domain also contains the fusion domain, which mediates fusion of the viral envelope with the plasma membrane of the host cell (Freed *et al*, 1990). Mutations within this region, can alter fusion in a cell-type specific manner (Bergeron *et al*, 1992).

The cytoplasmic domain of gp41 associates with the matrix protein (MA) $p17^{gag}$. Deletions in the gp41 cytoplasmic tail block incorporation of the envelope glycoproteins into virions, a block that is mediated by p17 (Freed and Martin, 1996).

1.3.10. Virus Entry into the Host Cell

Identification of the CD4 Molecule as the Cellular Receptor for HIV-1

The selective tropism displayed by HIV-1 for the subset of T-lymphocytes expressing the surface glycoprotein CD4, and the selective depletion of this cell-type in patients presenting with AIDS, indicated that the CD4 molecule was the cellular receptor used by HIV-1 for entry into the cell (Klatzman *et al*, 1984a).

Dalgleish *et al*, 1984, identified receptor-positive cells by assaying for the induction of syncytia following infection, and assaying the susceptibility of different cell types to infection by vesicular stomatitis virus (VSV) pseudotypes carrying the envelope molecules from HIV-1. Monoclonal antibodies directed at the CD4 molecule were found to block infection by HIV-1, and to block infection with VSV pseudotypes, and in addition, to block the formation of syncytia in HIV-1 infected cells (Dalgleish *et al*, 1984, Klatzmann *et al*, 1984b).

Further confirmation that the CD4 is the receptor for HIV-1 has come from the work of Maddon *et al*, 1984, who transformed various human and murine cell lines of lymphoid and non-lymphoid origin, with a cDNA clone of the human CD4 gene. The expression of the CD4 gene on the majority of human cell lines rendered these susceptible to infection by HIV-1. The introduction of the human CD4 gene into murine cells, however, did not render these cells permissive for HIV-1 infection, indicating the requirement for an additional factor present in human cells but absent from murine cells.

The CD4 molecule is found on a subset of T-lymphocytes (those of the helper subset), and on cells of the monocyte-macrophage lineage (Sattentau *et al*, 1989; Maddon *et al*, 1984).

Identification of Second Receptors for HIV-1 Entry into CD4+ T-Cells

The receptor molecule CC-CKR-5 has been identified as a second receptor required for the entry into CD4+ T-cells of certain strains of HIV-1 (Deng *et al*, 1996; Dragic *et al*, 1996). Co-expression of CC-CKR-5 and CD4+ on cells previously non-permissive for HIV-1 infection, enables infection by HIV-1 strains that have a macrophage-tropic, non-syncytium-inducing phenotype (which will be described below). The CC-CKR-5 gene is expressed in T-cell lines HUT78 and PM1, T-cells and monocyte/macrophages from peripheral blood (Deng *et al*, 1996).

Another receptor molecule, fusin (also called CKR4, LESTR or HUMSTR), is

required along with CD4, for the entry of T-cell-line adapted strains of HIV-1 into T-cells (Feng *et al*, 1996).

Mechanism of Viral Entry

Enveloped viruses enter the host cell by fusion of the viral envelope with the plasma membrane of the host cell, leading to the release of the nucleocapsid into the cytoplasm. With some enveloped viruses such as the Sendai virus and the paramyxoviruses, fusion occurs at the cell surface by direct fusion of the virus envelope and the cell membrane. Alternatively, with other viruses, such as the Semliki Forest virus, influenza type A virus, mouse mammary tumour virus and vesicular stomatitis virus, the virion may first undergo receptor-mediated endocytosis into acidic vesicles, where the envelope glycoprotein undergoes a requisite low-pH-dependent conformational change that facilitates fusion with the endosomal membrane (reviewed by White, 1990).

Viruses that enter the cell by receptor-mediated endocytosis often show a pH-dependent membrane fusion process, whereby, a low pH causes a conformational change in the envelope protein which activates its fusion potential. Virus entry may be perturbed by treatment of the cells with lysosomotropic agents such as ammonium chloride and chloroquine, which neutralise the endosomal compartments (McClure *et al*, 1988). In contrast, viruses that enter the cell by direct fusion at the plasma membrane show a pH-independent entry into the cell.

There is evidence to suggest that HIV-1 can enter the host cell by both direct fusion at the plasma membrane and by receptor-mediated endocytosis. For example, Maddon *et al*, 1984, observed a 95% inhibition of viral infection in cells treated with ammonium chloride, consistent with a mechanism of viral entry into the cell by endocytosis and low-pH-dependent fusion. However, others have shown that HIV-1 infection is not affected by treatment of the cells with lysosomotropic compounds (McClure *et al*, 1988). In addition, Pelchem-Matthews *et al*, 1995, have shown that inhibiting endocytosis of CD4 has no effect on the rate of HIV-1 infection.

Viruses that show entry into the cell by direct fusion at the plasma membrane, such

as, Sendai virus, lentiviruses, paramyxoviruses, and some herpesviruses, exhibit a syncytium-inducing property, whereby infected cell and non-infected cells fuse through the interactionbetween the viral envelope expressed on infected cells and the receptor molecule expressed the uninfected cells. In contrast, viruses that undergo fusion within endocytotic vesicles will not be syncytium-inducing.

Some isolates of HIV-1 form syncytia on infection of peripheral blood mononuclear cells, or T-cell lines. In contrast, some isolates are unable to form syncytia (Tersmette *et al*, 1988; Fenyo *et al*, 1989). The ability to induce syncytia *in vitro* is a property of the *env* gene: only a small number of amino acid changes are required to change a viral clone from a syncytium-inducing type to a non-syncytium-inducing virus (section 1.4). It is tempting to speculate that the envelope protein of syncytium-inducing isolates allows entry into the cell, via attachment to the appropriate second receptor by *direct fusion* at the cell surface. In contrast, the envelope of non-syncytium-inducing isolates may requires *endocytosis* for viral entry, in a process requiring the second receptor. This is discussed further in section 1.4.

1.4. Biological Properties Displayed by HIV-1 *In Vitro*

Three biological properties displayed by HIV-1 *in vitro* have been used to classify HIV-1 isolates:

- **the replication rate**: isolates may be classified as *rapid/high*, or *slow/low*, according to the rate and level of virus production (Fenyo *et al*, 1988, 1989; Asjo *et al*, 1986).
- **the cytopathic effect**: isolates may be classified as *syncytium-inducing* (SI), or *non-syncytium inducing* (NSI), dependent upon the ability of the virus to induce multinucleated cells, or syncytia (Tersmette *et al*, 1988).
- **the cellular host range**: HIV-1 isolates are often described as either *monocytotropic*, or *T-cell line tropic*, according to the ability of the virus to infect cells of the monocyte-macrophage lineage, or CD4+ transformed cell lines (Schiutemaker *et al*, 1991; Cheng-Mayer *et al*, 1989; Connor *et al*, 1994).

Variation in these properties is found between isolates taken from different individuals, and in isolates taken at different time points from the same individual (Cheng-Mayer *et al*, 1988, Tersmette *et al*, 1988). The analysis of the viruses present in individual isolates, examined through the production of biological or molecular clones, reveals that isolates frequently contain a multitude of variants, each of which may show distinct biological properties (Sakai *et al*, 1988, Schuitemaker *et al*, 1992, Connor *et al*, 1993a).

1.4.1. The Replication Rate

The distinction between slow/low and rapid/high isolates was first shown by Asjo *et al*, 1986, who examined the growth properties of primary HIV-1 isolates by co-cultivation of patient peripheral blood mononuclear cells (PBMCs) with donor PBMCs. The term rapid/high was assigned to isolates able to replicate efficiently in PBMCs. Conversely, slow/low isolates were defined as those only able to yield low amounts of virus after a prolonged time in culture.

Rapid/high and slow/low isolates can also be distinguished by their ability to induce syncytia and their transmission to CD4+ cell lines. Rapid/high isolates are generally found to induce syncytia and replicate in CD4+ cell lines, whereas slow/low isolates, in general, do not induce syncytia (Fenyo *et al*, 1988, 1989).

Although this classification system has facilitated the description of viral isolates, the differences in replication rate between the rapid/high and slow/low isolates, in reality, form a continuum, with no distinct cut-off between the two groups (Fenyo *et al*, 1989).

1.4.2. The Cytopathic Effect

There are two main cytopathic effects of HIV infection that can be observed in cell cultures of HIV-1 isolates. The first is the formation of syncytia, or multinucleated giant cells, caused by cell fusion (Fenyo *et al*, 1988, 1989; Tersmette *et al*, 1988). The second is single-cell death in the absence of syncytia, which is believed to be due to pycnosis and degradation (Fenyo *et al*, 1988 and 1989).

The induction of syncytia, and associated 'cell-ballooning,' which are readily observed under the light microscope, is frequently used to classify viral isolates into two groups: syncytium-inducing (SI), or non-syncytium-inducing (NSI) (Tersmette *et al*, 1988; Fenyo *et al*, 1989). The degree and number of syncytia produced by different isolates may vary but appears to be independent of the replication rate (Tersmette *et al*, 1989b; De Jong *et al*, 1993a).

In order to standardise measurements, a definition of syncytium-inducing capacity based upon the induction of syncytia in the T-cell line MT2, has been proposed (Koot *et al*, 1992). In this classification scheme, isolates are considered as syncytium-inducing if syncytia are observed upon infection of MT2 cells, and non-syncytium-inducing if no syncytia are observed.

1.4.3. The Cellular Host Range

HIV-1 shows a requirement, or a strong preference, for replication in cells expressing the CD4 receptor molecule (Klatzmann *et al*, 1984a). The most commonly used experimental cell culture systems are peripheral blood mononuclear cells (PBMCs) obtained from seronegative donors. PBMCs may be fractionated to obtain cultures of CD4+ T-lymphocytes, or cultures of monocyte-derived macrophages, both of which are permissible to HIV-1 infection (Schuitemaker *et al*, 1991; Cheng-Mayer *et al*, 1988). In addition, CD4-positive transformed cell lines of lymphoid or monocytoid origin are widely used (McKeating *et al*, 1989; Tersmette *et al*, 1988). Specific human cell lines transfected with the human CD4 molecule, such as HeLa CD4, are also permissive for HIV infection (Chesebro *et al*, 1991).

HIV-1 isolates show a differing ability to replicate in these different CD4+ cell types, which allows a measure of the cellular host-range, or 'cell tropism'. In general, most HIV-1 isolates grow in cultures of PBMCs and CD4-positive lymphocytes (Asjo *et al*, 1986; Schuitemaker *et al*, 1991). Marked variation, however, is seen in the ability of different isolates to replicate in CD4-positive transformed cell lines and monocyte-derived macrophages (Gendelman *et al*, 1990).

In general, HIV-1 isolates that are capable of growth in cultures of monocyte-derived macrophages fail to infect transformed T cell lines. These isolates are described variously in the literature, as monocytotropic, MT-tropic (for macrophage and T-lymphocyte), or of broad host-range, or dual tropic (Moore and Ho, 1995).

Conversely, isolates capable of growth in CD4-positive cell lines are generally unable to replicate in monocyte-derived macrophage cultures. These isolates are referred to as T-cell-line tropic, or occasionally as simply T-tropic, or of narrow-host range - referring to their loss of ability to grow in macrophages.

However, many researchers have found that all primary isolates can replicate in monocyte-derived macrophages (Cheng-Mayer *et al*, 1988; Valentin *et al*, 1994; Connor *et al*, 1993b, 1994, Fiore *et al*, 1994). Valentin *et al*, 1994, who found that even established T-cell-line tropic strains, such as HIV-1_{IIIB} and HIV-1_{SF2}, could replicate in macrophages, have suggested that it is the method of isolation and culturing that explains the discrepancy between different laboratories in the ability to propagate isolates in monocyte-derived macrophages.

The relative ability of different isolates to replicate in CD4+ T-lymphocytes compared to monocyte-derived macrophages varies widely. For example, the isolate HIV-1_{IIIB} shows a 1000-fold preferential ability to infect T-lymphocytes compared to infection of monocyte-derived macrophages. In contrast, the HIV-1 isolates HIV-1_{ADA} and HIV-1_{BAL}, show a 70-fold, and 10-fold increased infectivity, respectively, for monocyte-derived macrophages compared to T-lymphocytes. However, the prototype macrophage isolates, HIV-1_{SF162} and HIV-1_{JR-FL}, show a 10-fold greater infectivity for T-lymphocytes compared to monocyte-derived macrophages (Collin *et al*, 1994).

There is a correlation between the three biological properties displayed by HIV-1 isolates, leading to a broad classification of viral strains into two groups: firstly, slow/low replicating isolates which have a non-syncytium-inducing phenotype, and are unable to grow in transformed cell-lines, but are capable of growth in monocyte-derived macrophages; secondly, rapid/high replicating isolates which are generally syncytium-inducing, capable of growth in transformed cell lines but not in monocyte-

derived macrophages. There are, however, exceptions to this classification, in particular, there have been reports of rapidly replicating, syncytium-inducing isolates capable of infecting macrophages (Collman *et al*, 1992; Connor *et al*, 1994; Valentin *et al*, 1994).

1.4.4. The Relationship between the Biological Phenotype of HIV-1 Isolates and Disease Status of the Host

Early studies with isolates of HIV-1 showed that the biological characteristics of an isolate are closely related to the severity of the disease in the infected individual. In cross-sectional studies, isolates from patients with a more severe disease were often found to have a high replicative capacity, and were able to grow in transformed CD4+ T-cell lines. In contrast, isolates with lower replicative capacity were isolated from patients in the asymptomatic stage of disease, and these generally could not grow in transformed CD4+ T-cell lines (Asjo *et al*, 1986; Fenyo *et al*, 1988; Tersmette *et al*, 1988).

In longitudinal studies, patients who remained asymptomatic during the time period studied were found to have low replicating, non-syncytium-inducing isolates that could not be transmitted to T-cell lines. Patients who progressed slowly to AIDS had higher replicating, non-syncytium-inducing isolates that could not be transmitted to T-cell lines. In contrast, patients that showed a rapid progression to AIDS had high replicating, syncytium-inducing variants that could be transmitted to T-cell lines (Tersmette *et al*, 1989b; Connor *et al*, 1993b; Oka *et al*, 1994).

Subsequent studies have shown that the emergence of T-cell line tropic, syncytium-inducing variants, is associated temporally with a rapid CD4+ T-cell decline and the development of AIDS in 50% of patients (Koot *et al*, 1993). It has been speculated that the emergence of these more cytopathic variants is the cause of the decline in CD4 cell count and leads to the progression of AIDS; in other words, that a more 'virulent' strain of HIV-1 emerges during the course of infection (Tersmette *et al*, 1989b, Schuitemaker *et al*, 1992). However, whether the emergence of these more

cytopathic variants is a cause or consequence of immune system breakdown is not known. In particular, 50 % of patients that progress to AIDS do not have syncytium-inducing (SI) variants, and SI variants can often be found at times much earlier than the time of progression to AIDS (Tersmette *et al*, 1988; Fiore *et al*, 1994). In addition, isolates that are cytopathic *in vitro* may not be cytopathic *in vivo* (Mosier *et al*, 1993).

Whilst many of the early studies correlating viral phenotype *in vitro* and disease progression have highlighted the importance of syncytium-inducing, T-cell line tropic variants, other more recent studies have shown the importance of macrophage-tropic isolates in HIV-1 infection (Connor *et al*, 1993b; Gartner and Popovic, 1990; Cheng-Mayer *et al*, 1988, 1989).

Many authors have shown that isolates taken in later stages of disease retain their ability to replicate monocytes-derived macrophages (Gartner and Popovic, 1990; Connor *et al*, 1993b, 1994), and may show an increased replicative ability in this cell type compared to earlier isolates (Connor *et al*, 1993b, 1994, Cheng-Mayer *et al*, 1988). These later isolates, in addition, frequently acquire the ability to form syncytia and the ability to grow in T-cell lines, in agreement with earlier studies (Tersmette *et al*, 1989a, 1989b; Schuitemaker *et al*, 1992).

Macrophage-tropic isolates can be found in the peripheral blood at all stages of disease, despite the emergence of variants with a greater ability to grow in T cell lines, which generally lose their ability to grow in macrophages (Schuitemaker *et al*, 1991, 1992). It has been suggested that infection of macrophages with HIV-1 enables the persistence of HIV-1 and its spread outside the peripheral blood. Indeed, the macrophage appears to be the most important cell harbouring HIV-1 in tissues of non-lymphoid origin (Koenig *et al*, 1986; Wiley *et al*, 1986). Viruses isolated from tissues appear to have a higher ability to grow in cultures of monocyte-derived macrophages than do peripheral blood-derived isolates (Schuitemaker *et al*, 1992; Cheng-Mayer *et al*, 1989).

Isolates taken around the time of primary infection are frequently macrophage-tropic

(Schuitemaker *et al*, 1992; Zhu *et al*, 1993; Fiore *et al*, 1994; Kliks *et al*, 1994). It has been suggested that following primary infection, for both sexual and vertical transmission, macrophages present in the mucosa and placental tissues may be the first target cells encountering HIV-1 (Miedema, 1992; Braathen *et al*, 1987). However, it has been suggested that SI variants are suppressed by the immune system during the first years of asymptomatic infection and this may delay the pathogenic process of infection (Schuitemaker *et al*, 1992). In a second study, individuals in which variants with a rapid/high phenotype were present on primary infection, showed a more rapid CD4 decline than those infected with slow/low viruses (Fiore *et al*, 1994). In this study also, all isolates were capable of replication in monocyte-derived macrophages.

The macrophage is also an important cellular target of ungulate lentiviral infections - specifically Visna virus of sheep, and caprine arthritis encephalitis virus of goats. Disease is characterised by virus-infected macrophages in several target organs, such as the lung, joints and CNS, where they initiate inflammatory responses which result in extensive organ damage (Haase, 1986).

Insight into the possible importance of the infection of macrophages in the pathogenesis of HIV-1 has come from work with the experimental infection of chimpanzees with HIV-1 (Schuitemaker *et al*, 1993). It is known that chimpanzees may be infected long-term with HIV-1, but do not develop clinical symptoms. In the infected animals systemic immune dysfunction does not occur; antigen-presenting function remains intact, T-cells have normal proliferative capacity with no evidence of programmed cell death, which are often cited as evidence of immune dysfunction in infected humans. Low viral loads can be demonstrated in the peripheral blood at levels of 1 in 10^3 to 10^4 infected cells. Virus may be isolated from the peripheral blood of these animals; however, these isolates show an exclusive tropism for chimpanzee peripheral blood lymphocytes, and are unable to replicate on either chimpanzee or human monocytes. The lack of ability of HIV-1 to infect chimpanzee monocytes is believed to result in an absence of immune dysfunction, and the ability to control viral replication. However, these experiments were conducted with the T-cell line adapted isolate HIV-1_{IIIB} (Shaw *et al*, 1984), which is unable to replicate in

monocyte-macrophages, or to very low levels (Valentin *et al*, 1994).

1.4.5. Mapping the Determinants of Viral Phenotype

Studies with recombinant proviral clones generated by swapping different subgenomic regions of clones with distinct phenotypes, have shown that the *principal determinant* of viral replication rate, cytopathic effect and cellular host-range, resides in the *env* gene (O'Brien *et al*, 1990; Helseth *et al*, 1990b; York-Higgins *et al*, 1990; Cheng-Mayer *et al*, 1991; Chesebro *et al*, 1991; Westervelt *et al*, 1991; Shioda *et al*, 1991; Groenink *et al*, 1992; Cann *et al*, 1992; Fujita *et al*, 1992; Sullivan *et al*, 1993; Stamatatos *et al*, 1994).

In many cases, only a few amino acid substitutions within the V3 loop were sufficient to change the phenotype of resultant clones. Amino acid substitutions in the V3 loop can alter the syncytium-inducing capacity of molecular clones (De Jong *et al*, 1993a, 1993b), and the cellular tropism (Takeuchi *et al*, 1991; Hwang *et al*, 1991; Shimizu *et al*, 1994; Stamatatos *et al*, 1994; Willey *et al*, 1994).

However, other regions of the *env* gene have been found to influence the phenotype. For example, the V1 and V2 regions act in combination with the V3 loop to alter syncytium-inducing ability and replication rate (Andeweg *et al*, 1993; Groenink *et al*, 1992, 1993; Sullivan *et al*, 1993; Koito *et al*, 1994). In addition, the cellular tropism can be altered by single mutations in the V1 region (Boyd *et al*, 1993).

Other genes, or genomic regions, also affect virus phenotype. For example, the replication rate in different T-cell lines may be altered by changing the sequences of the Rev-responsive element (Dayton *et al*, 1993), or by mutations in *tat* gene (Cheng-Mayer *et al*, 1991). The cellular host-range may be narrowed by the absence of a *vpu* or *vif* gene, which are required in conjunction with specific *env* sequences for replication in monocyte-derived macrophage cultures (Westervelt *et al*, 1992b, Du *et al*, 1993). However, the LTR does not appear to be involved with *in vitro* phenotype in HIV-1 (Hirsch *et al*, 1990; Pomerantz *et al*, 1991)

1.4.6. The Relationship between the Amino Acid Sequence of the V3 Loop and the *In Vitro* Viral Phenotype

Whilst studies with recombinant clones have demonstrated the influence of specific amino acids or regions of the *env* gene on the phenotype, each used artificially-constructed recombinant viruses; the effect of these changes may be influenced by the background of the virus used. Perhaps more relevant to determining the phenotype of isolates obtained from infected patients are studies of panels of HIV-1 isolates which aim to correlate specific sequence-motifs, or amino acid substitutions, with a given phenotype (Fouchier *et al*, 1992; Milich *et al*, 1993).

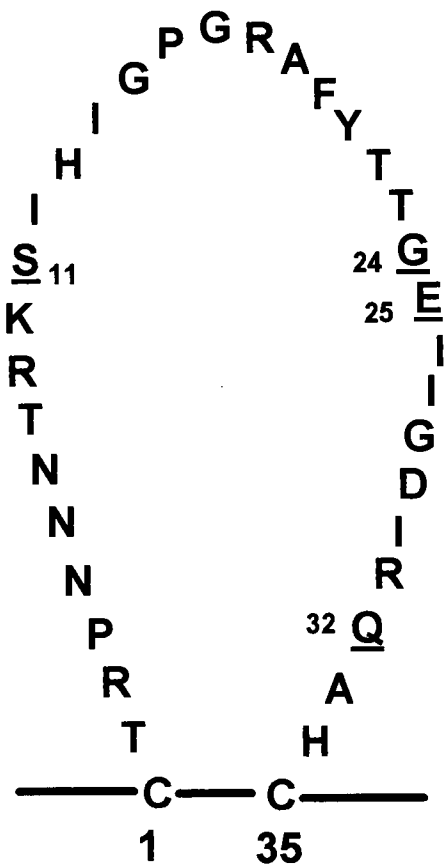
An early indication that T-cell line tropic and macrophage-tropic viruses may constitute two different groups according to the amino acid sequence of their V3 loop was provided by Chesebro *et al*, 1992. In this study, it was observed that the V3 loop of macrophage-tropic isolates showed a more restricted set of mutations as compared to T-cell-line tropic variants.

Fouchier *et al*, 1992, showed through comparisons of the V3 loop amino acid sequence from a panel of sequential biological clones that there is a correlation between the acquisition of basic amino acids (K and R) at positions 11 and 25 in the V3 loop and the transition from an NSI to an SI phenotype (figure 1.5). The presence of a basic amino acid at one or both of these sites has been shown experimentally to confer an SI-phenotype on the resultant viral clone (De Jong *et al*, 1993b).

In a more extensive study, Milich *et al*, 1993, examined a large data set consisting of well-characterised HIV-1 isolates for which the sequence of the V3 loop was known. A correlation was found between both the charge of the V3 loop - in terms of the presence of basic amino acids at specific residues - and the sequence heterogeneity of the loop - and the *in vitro* phenotype. All T-cell-line tropic viruses had basic amino acids (K and R) in at least one of the four positions: 11, 24, 25 32, which was generally associated with a change away from an acidic residue (E/D) at position 25. In contrast, no macrophage-tropic isolates had a basic substitution at any of these positions. In addition, isolates with basic amino acid substitutions, showed a two-fold greater number of amino acid differences from the North American

Figure 1.5

The V3 loop of HIV-1
(subtype B consensus sequence)



Residues underlined are the sites whose charge is important in determining the syncytium-inducing ability and cellular host-range (see text for details).

(subtype B) V3 loop consensus sequence (Myers *et al*, 1991).

A further analysis along the same lines by Donaldson *et al*, 1994b, showed that in combination the net charge of the V3 loop and number of amino acid differences from the subtype B consensus sequence produces an almost complete separation of published variants into two groups. The first contains the NSI/macrophage-tropic isolates, which have a low net charge and greater similarity to the subtype B consensus. The second contains the SI/T-cell-line tropic variants, which are of higher charge and greater heterogeneity.

1.4.7 The Role of Second Receptors in HIV-1 Tropism

The CC-CKR-5 molecule is a co-receptor with CD4, allowing fusion and entry of NSI, macrophage and T-lymphocyte-tropic strains into T-lymphocytes (T-cells) (Deng *et al*, 1996; Dragic *et al*, 1996). In contrast, the receptor molecule, fusin, is required for fusion and entry of T-cell line-tropic, SI strains (Feng *et al*, 1996).

A model for receptor usage and cell tropism has been proposed by Deng *et al*, 1996. The envelope protein of isolates found early in infection, that is, NSI, T-cell and macrophage-tropic, may have a high binding affinity for CC-CKR-5 which is expressed in macrophages and T-cells. In contrast, the switch to a T-cell-line tropic virus, seen during the course of infection, or following prolonged passage *in vitro*, may be due to changes in envelope configuration, allowing the use of fusin, as well as, or instead of CC-CKR-5. Fusin is presumed to be expressed on T-cells and T-cell lines.

1.4.8 Possible Role of the V3 Loop in Viral Life-Cycle

Whilst there is convincing evidence demonstrating the role of the V3 loop in determining the cellular host range and syncytium-inducing ability of HIV-1, the mechanism and the stage in the viral life-cycle at which the V3 loop exerts its effect(s) is currently not known.

The V3 loop resides outside the CD4-binding site, and the hydrophobic fusion peptide

which resides in gp41. Antibodies to the V3 loop which act to prevent viral infection, act at a step after the binding of the virion to CD4 (Page *et al*, 1992). This suggests a role for the V3 loop in a post-CD4-binding step of viral entry.

It is known that the envelope protein undergoes a conformational change following binding to CD4 (Sattentau and Moore, 1991). Upon binding, the V3 loop adopts a more exposed position on the envelope complex (Sattentau *et al*, 1995). This change in conformation, which may be influenced by the structure and charge of the V3 loop, is believed to trigger or activate the fusion potential of the virus.

With the recent discovery of different co-receptors for entry of T-cell line SI strains, and NSI macrophage strains, it has been speculated that a post-CD4 binding conformational change in the HIV-1 envelope may expose a co-receptor binding site. This co-receptor binding site may contain the V3 loop or be influenced by its conformation (Deng *et al*, 1996; Dragic *et al*, 1996). Binding to the appropriate co-receptor, may induce further conformational changes activating the fusion domain in gp41.

1.5. The Genetic Variation of HIV

1.5.1. The Quasispecies

One of the most important properties displayed by HIV is its genetic variability. Indeed, it is now recognised that most viral species, especially the RNA viruses, which display a high mutation rate, consist of a genetically complex population that comprises multiple variants.

The term *quasispecies* has often been used to indicate the heterogeneity of a virus population. A quasispecies, however, is more than an heterogeneous population of virus genomes, but a mathematically-defined distribution of mutants generated by a mutation-selection process (Eigen *et al*, 1988). The term was originally used in theoretical work on the origin of life, to describe a cluster of closely related molecular species, produced by errors in the self-replication of the molecules. For

example, in the situation where there are no errors during self-replication, and an array of mutants with different replication rates, a homogeneous population will arise by Darwinian selection, consisting of the fastest-replicating variant. However, when the error-rate is high - such as for RNA viruses - a distribution of mutants is generated (the quasispecies), which will eventually reach an equilibrium. At the population equilibrium, a consensus (or 'master') sequence is produced, around which the mutants are distributed. This has been experimentally documented for *in vitro* systems (Domingo *et al*, 1978).

The importance of the quasispecies is that the target of selection is the mutant distribution, rather than a single sequence. In the case of a population of RNA viruses, many different phenotypic variants may co-exist in the quasispecies. These may then be the target of selection: for example - resistant forms that are no longer recognised by the host's immune system - as has been found for influenza A virus (Fitch *et al*, 1991), and proposed for lentiviral infections (Carpenter *et al*, 1987; Albert *et al*, 1990; Burns *et al*, 1993), and for the development of drug resistant mutants - such as AZT resistance in HIV-1 infection (Larder *et al*, 1989). There may be the potential for the selection of more virulent or pathogenic forms, both within the individual, as has been postulated for HIV-1 and -2 (Tersmette and Schuitemaker, 1993, Nowak and May, 1993), and within the infected population as a whole - such as has occurred in myxoma virus infection of rabbits (Fenner and Kerr, 1994).

Demonstration of Genetic Variability of HIV-1

HIV-1 shows considerable variation across the whole genome. However, the rate of substitution, is not uniform, but is particularly high in the *env* gene and the regulatory genes *tat*, *rev*, *nef*, *vif* and *vpr*. The *gag* gene encoding the virion structural proteins, and the *pol* gene encoding the viral enzymes, show an approximately ten-fold lower rate of nucleotide substitution (Li *et al*, 1988).

Variation in the *env* gene, which shows the highest rate of substitution, is not uniform. Early studies, comparing *env* gene sequences from different isolates, revealed the presence of five *hypervariable* regions, designated V1 to V5, which show less than 25% conservation of amino acid sequence (Modrow *et al*, 1987,

Starcich *et al*, 1986). These hypervariable regions map to the surface glycoprotein (gp120) portion of the *env* gene (figures 1.4 and 1.6). Interspersed between the hypervariable regions are more conserved regions, designated the conserved regions C1 to C5. The transmembrane protein (gp41) does not contain any defined hypervariable regions, but does show extensive variation between different HIV-1 sequences (Myers *et al*, 1991).

Source of Variation

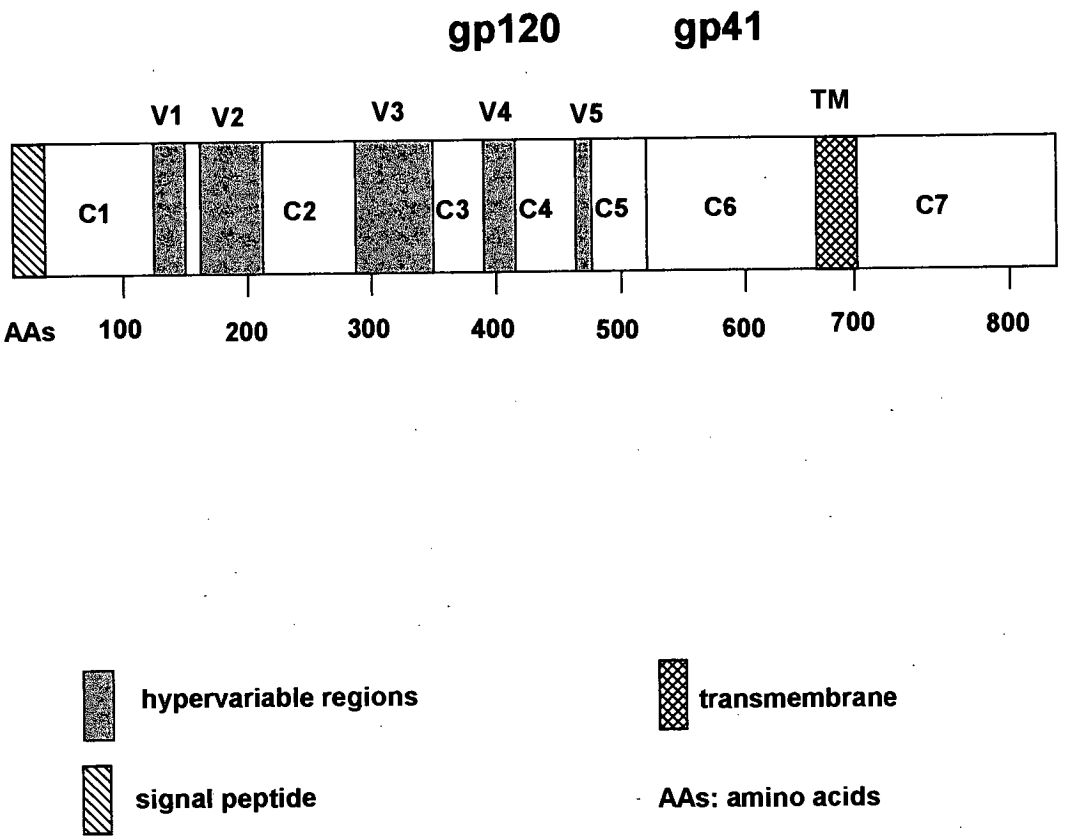
A major source of the genetic variation of HIV is from the error-prone viral polymerase enzyme - reverse transcriptase. This enzyme lacks a 5'->3' exonuclease or 'proof-reading' ability, and therefore cannot repair nucleotide mismatches introduced into the sequence during the synthesis of the viral DNA. The frequency of the introduction of point-mutations into the genome has been estimated as $2x10^{-3}$ to $5x10^{-3}$ per nucleotide per replication cycle (Preston *et al*, 1988, Roberts *et al*, 1988). However, Mansky and Temin, 1995, found an '*in vivo*' error rate using an HIV-1 based shuttle vector replicating in HeLa cells, to be $3.4x10^{-5}$, 100-fold lower than the *in vitro* rate. This figure is, however, 3 to 7 times higher than those reported for other retroviruses.

Other mutational processes have been described for HIV-1. For example, the process of *hypermutation*, has been described, whereby a predominance of G to A transitions occurs (Goodenow *et al*, 1989). The mechanism for this is unclear, but appears to be due to a property of the reverse transcriptase. In addition, during the viral life cycle, the provirus is transcribed by the cellular RNA polymerase II, which is another error-prone polymerase, providing a further source of mutation.

Recombination between non-identical viral genomes present in the same virion can occur under experimental conditions. During reverse-transcription, template-switching by the reverse-transcriptase enzyme may result in the generation of a 'hybrid' double-stranded DNA version that is recombinant between the two RNA molecules (Hu and Temin, 1990). There is evidence of recombination occurring during the course of natural infection with HIV-1 (Clavel *et al*, 1989, Howell *et al*, 1991, Simmonds *et al*, 1991, Diaz *et al*, 1995).

Figure 1.6

The Conserved and Variable Regions of gp160



The Different Levels at which the Genetic Variation may be Examined

The genetic variation and evolution of HIV can be examined at several levels. The genetic differences between the human and primate immunodeficiency viruses can be used to determine their evolutionary relationships - which may shed light on the origin of HIV-1 and HIV-2 (Myers *et al*, 1992; and Eigen and Nieselt-Struwe, 1990). Comparisons of the rates of amino acid changes in human and simian immunodeficiency viruses, have been used to examine the evolution of viral virulence (Shpaer and Mullins, 1993, Nowak and May, 1993).

The use of genetic relatedness of HIV sequences (presented as a phylogenetic tree), can be used to trace the spread of HIV - globally (Louwagie *et al*, 1993; Charneau *et al*, 1994), within a community (Holmes *et al*, 1995; Ou *et al*, 1992), or indeed, within the different tissue compartments of an infected individual (Donaldson *et al*, 1994b; Korber *et al*, 1994). This type of analysis is known as molecular epidemiology.

1.5.2. The HIV Subtypes

HIV-1 Subtypes

In order to characterise the global diversity of HIV-1, isolates from different parts of the world have been examined. Extensive molecular epidemiology studies, using both sequence information and genetic heteroduplex analysis, have shown that isolates of HIV-1 fall into two genetic clusters, which appear to be the result of two independent introductions of primate lentiviruses into the human population (Louwagie *et al*, 1993; Myers *et al*, 1991; Delwart *et al*, 1994).

The first cluster has been named group M, for (major group) (Charneau *et al*, 1994) and contains the majority of HIV-1 isolates found world-wide. The second cluster which has been named group O (for outgroup) was discovered recently, with a small number of cases from infected patients in Cameroon and more recently of a French woman who had never travelled outside Europe, nor identified a sexual partner from Africa, nor had received any known blood transfusion. (Charneau *et al*, 1994).

Group M is further divided into 8 distinct genetic lineages (or *clades*), on the basis of both *gag* and full-length *env* sequence data (Louwagie *et al*, 1993), differing by 10 to 13% in their nucleotide sequence. The genetic lineages define 8 HIV-1 subtypes, denoted A to H. A ninth subtype I, has recently been recognised, from patients in Cyprus (Kostrikis *et al*, 1995).

The global distribution of the subtypes is complex, with some subtypes showing a restricted distribution, whilst others are distributed in many continents, and as a consequence many locales, such as Central Africa, have multiple subtypes (table , below).

Table 1. Summary of the global distribution of HIV-1 subtypes from group M

Subtype	Predominant distribution
A	Central and Western Africa
B	North America and Europe, Brazil
C	South and Central Africa
D	Central Africa
E	Northern Thailand, South and Central Africa
F	Brazil, Rumania, Zaire
G	Gabon and Zaire
H	Gabon
I	Cyprus

The subtypes have been used to study the epidemiology of HIV-1 infection world-wide (reviewed by Leigh Brown and Holmes, 1994). Each subtype has been shown to eventually cause AIDS and to date, there is no evidence to suggest a difference in

the rate of pathogenesis of the different subtypes.

In terms of vaccine design, extensive serological research is being undertaken with the eventual aim of quantifying the number of different epitopes required to produce a vaccine capable of counteracting infection with the diverse range of variants that may be encountered in the field (Gao *et al*, 1994 & 1996; Moore *et al*, 1994c & 1996). In a study to determine the extent of neutralisation serotypes, Moore *et al*, 1996, found that although the different neutralisation serotypes did not correspond to the genetic subtypes, some sera were found to neutralise across clades. This raises hope for the development of a broadly effective vaccine.

HIV-2 Subtypes

The extent of genetic diversity of HIV-2 has been studied by Gao *et al*, 1994, who obtained viral sequences from West African HIV-2 sero-positive asymptomatic and AIDS patients from whom virus isolation had repeatedly been unsuccessful. Phylogenetic analysis using *gag*, *pol* and *env* sequence data, revealed five distinct genetic lineages, denoted, by analogy to HIV-1, subtypes A to E.

There are some preliminary indications that there may be significant differences in the biology of the different HIV-2 subtypes: viruses from type A were able to grow to high titre in cell culture, whereas all attempts to obtain isolates from type C, D and E, as well as the majority of subtype B failed (Gao *et al*, 1994). However, greater numbers of patients will be required to demonstrate whether the inability to isolate is a genuine phenomenon or a biased result due to the small number of patients studied.

1.5.3. The Evolution of HIV-1 During the Course of Natural Infection

The degree of genetic variation of the *env* gene of HIV-1 within an infected individual varies according to the disease status of the patient. Three phases of evolution: I, II and III, have been identified (McNearney *et al*, 1990 and 1992; Leigh Brown, 1991; Wolfs *et al*, 1992; Zhang *et al*, 1993; Zhu *et al*, 1993). These coincide

with the three clinical phases of HIV infection: primary infection, the asymptomatic phase and the symptomatic phase. It believed that the three phases of sequence evolution are a result of different selective processes acting upon the virus population at these time points.

Phase I: the sequence diversity observed upon primary infection, prior to the development of the anti-viral immune response. A number of researchers have found that there is a restricted diversity within the *env* gene on primary infection (Wolfs *et al*, 1992; McNearney *et al*, 1990; Ait-Khaled and Emery, 1993; Zhu *et al*, 1993; Zhang *et al*, 1993). This restricted diversity is seen following either one of the three main routes of transmission: sexual intercourse (Kuiken *et al*, 1993; Zhu *et al*, 1993), transmission via blood or blood products (Cichutek *et al*, 1991; McNearney *et al*, 1992; Zhang *et al*, 1993), and mother-to-child transmission (Wolinsky *et al*, 1992; Scarlatti *et al*, 1993; Mulder-Kampinga *et al*, 1993).

In the study of Zhang *et al*, 1993, no sequence variation was found in the V3 and V4 regions of the *env* gene within any of the infected patients. In addition, the amino acid sequences of the V3 loop were extremely similar in different patients, many of whom showed epidemiologically unlinked infections (Holmes *et al*, 1995). This was in marked contrast to the sequences of the *gag* gene, which showed diversity in a number of patients. A similar result was found by Zhu *et al*, 1993, although in this case, less similarity was found in the V3 loop sequence. Again, a greater diversity was found in the sequences of the *gag* genes, suggesting a constraint on the *env* gene diversification.

This restriction in diversity in the *env* gene, which is not observed for the *gag* gene, may be the consequence of a strong selective force acting during primary infection, or during transmission. The result of this force produces extremely homogeneous V3 loop sequences within the individual (Zhang *et al*, 1993) and similar sequences between patients (Kuiken *et al*, 1993; Zhang *et al*, 1993).

The majority of V3 loop sequences found at primary infection are very similar to those of macrophage-tropic isolates (Wolinsky *et al*, 1992; Mulder-Kampinga *et al*,

1993; Scarlatti *et al*, 1993; Zhu *et al*, 1993; Zhang *et al*, 1993). In accordance with this, many virus isolates taken at primary infection are macrophage-tropic and NSI (Schuitemaker *et al*, 1992; Zhu *et al*, 1993), but not exclusively so (Clark *et al*, 1991; Fiore *et al*, 1994).

It is conceivable that the macrophage is the most efficient cell at transmitting the virus, or that macrophages provide the primary target for virus infection after transmission to the new host. However, it is conceivable that the limited heterogeneity found at primary infection could be due to a sequence bottleneck occurring during transmission, with infection initiated by a very few random variants. However, if this were so, we would not expect to see such a high degree of diversity in *gag*.

Phase II: This is the phase of asymptomatic infection in which there is continual viral replication in the peripheral blood and lymph nodes, despite the presence of a strong anti-viral immune response (Zhang *et al*, 1991; Piatek *et al*, 1993; Michael *et al*, 1992). Several studies have demonstrated that in the peripheral blood at this time the virus population shows a high degree of genetic heterogeneity and, in addition, there is a rapid turnover of sequence variants with time (Hahn *et al*, 1986; Meyerhans *et al*, 1989; Howell *et al*, 1991; Simmonds *et al*, 1990a, 1991; Wolfs *et al*, 1990, 1992; Holmes *et al*, 1992; Cichutek *et al*, 1992; Kusumi *et al*, 1992; McNearney *et al*, 1992; Delwart *et al*, 1994; Lukashov *et al*, 1995).

Similar studies with experimentally-infected monkeys have shown that rapid sequence diversity arises over the course of infection (Burns and Desrosiers, 1991; Baier *et al*, 1991, Johnson *et al*, 1991).

Much of the variation in the V1 to V5 of the *env* gene results in amino acid changes at specific sites, and these are frequently found to alter the number of N-linked glycosylation sites. In addition, there is considerable change in the length of the hypervariable regions, although the length of the V3 loop is relatively conserved at 35 amino acids (Simmonds *et al*, 1990b, 1991; Holmes *et al*, 1992; Howell *et al*, 1991; Cichutek *et al*, 1992). These changes may represent variation that results in the

generation of antigenically-distinct variants.

Phylogenetic analyses have shown that the evolution of the *env* gene during the asymptomatic period is complex. In the hypervariable regions, often two or more distinct viral populations exist at any one time. These populations will often differ by a number of point mutations, and/or the length of the variable region (Howell *et al*, 1991; Simmonds *et al*, 1991; Holmes *et al*, 1992; Cichutek *et al*, 1991). The distinct populations evolve in parallel, so the sequence evolution does not follow a simple linear path where one variant is replaced by its descendants; distinct variants arising from the other population will also be present. This process of evolution is discussed in detail in chapter 5 of this thesis.

Amino acid substitutions in the V3 loop observed at this time are often at sites that alter antibody recognition when synthetic peptides representing the V3 loop are used to assay antibody-binding (Wolfs *et al*, 1991; Zwart *et al*, 1991). Indeed, Wolfs *et al*, 1991, have shown that the V3 sequence variants which emerge during the course of infection are accompanied by the emergence of a new population of V3-specific antibodies.

The argument that selection is a dominant force driving sequence evolution during this phase of infection can be made by determining the proportion of synonymous nucleotide substitutions (K_S) to non-synonymous substitutions (K_A), expressed as the K_S/K_A ratio (Simmonds *et al*, 1990b; Wolfs *et al*, 1992). In these studies, the K_S/K_A ratio of the V3 region, was found to be less than 1, which is indicative of positive selection. A similar predominance of non-synonymous substitutions was found by Burns and Desrosiers, 1991, who examined the evolution of SIV in experimentally-infected rhesus monkeys.

Together, these data strongly suggest that the variants within the *env* gene evolve under selection pressure from the immune system.

Phase III: This stage begins at the time when the immune system breaks down and the patient progresses to AIDS. There is an increase in viral burden at this time,

presumably due to the loss of constraint on viral replication from the immune system. HIV-1 infection, which during the asymptomatic phase is limited to the lymphoid tissues and peripheral blood, spreads to many other organs of the body, including tissues of non-lymphoid origin such as the brain, lung and muscle (Donaldson *et al*, 1994a). By analogy to phase I, which occurs before the immune response develops, it has been predicted that in the absence of immune control, the fastest replicating viruses will predominate, and this will reduce the variation observed (Nowak *et al*, 1990).

A limited variation was seen in the *env* gene of two infected neonatal patients who rapidly progressed to AIDS following blood transfusion (McNearney *et al*, 1990). However, in a study of four patients who died of AIDS-defining illness, a wide range of sequence variation was observed in the V3 loop of sequences taken from peripheral blood cells, lymph nodes and non-lymphoid organs (Donaldson *et al*, 1994b). This sequence diversity was equal to that seen in three asymptomatic patients, who died of non-AIDS-related causes.

1.5.4. Sequence Variation and Tissue Tropism *In Vivo*

HIV has been found to infect cells of both lymphoid and non-lymphoid origin - such as the central nervous system, lung, small and large intestine (Levy, 1993).

In the peripheral blood, the predominant cell type infected is the memory CD4+ lymphocyte (Schnittman *et al*, 1990), although others have reported the presence of HIV sequences in monocytes and dendritic cells, but these appear to be infected at a lower frequency than CD4 lymphocytes (Macatonia *et al*, 1990; Yamashita *et al*, 1994). In the lymph nodes HIV-1 is found in CD4+ lymphocytes and macrophages, and associated with follicular dendritic cells (FDCs) - where it is thought virions are trapped by cellular processes of the FDCs (Embretson *et al*, 1993; Pantaleo *et al*, 1993).

In contrast, HIV infection of non-lymphoid tissue appears to be mainly limited to tissue macrophages. For example, in the brain the most predominantly infected cell

is the macrophage and microglial cells (Koenig *et al*, 1986; Wiley *et al*, 1986). Infection in the lung is also associated with macrophages, although there is some evidence of lymphocytic infiltration of lung tissue in AIDS patients which may harbour HIV (Donaldson *et al*, 1994b; Bell *et al*, 1993).

Based upon sequence analysis, a number of researchers have found that the non-lymphoid tissues harbour different viral populations than the peripheral blood or lymphoid tissue (Epstein *et al*, 1991; Pang *et al*, 1991; Ball *et al*, 1994; Itescu *et al*, 1994).

Some have addressed the question of whether there is a specific adaptation of HIV variants for different cell types, and whether these sequence differences reflect the differences found between *in vitro* macrophage or T-cell line tropic variants. This approach was taken by Donaldson *et al*, 1994b, who examined proviral DNA sequences from a number of organs, both lymphoid- and non-lymphoid, taken at post mortem from three patients who died while in the asymptomatic stage of infection, and four patients who died with AIDS-defining illness. In patients dying of AIDS, there were distinct populations of HIV in different organs, but there was no consistent segregation of variants with specific V3 sequences into different cell types. For example, variants with the same V3 loop were found in proviral sequences recovered from the brain and lymphoid tissue. In other cases, there were distinct populations at different sites within a single organ, suggesting a random component in the distribution of *env* sequences *in vivo*. The inferred phenotype of the V3 sequences of variants from lymphoid tissue and non-lymphoid tissue was NSI for all but 3 sequences, in both the symptomatic and asymptomatic individuals. Indeed, the V3 loop was found to be relatively invariant compared to the regions flanking the V3 loop and in the V4 region (an observation made by Zhang *et al*, 1993).

A similar result was found in the study of Korber *et al*, 1994 who found that sequences derived from the brain of six AIDS patients taken at postmortem showed a low net charge, similar to that described for macrophage tropic isolates. There was significantly less sequence variation between sequences obtained from the brain

compared to those derived from peripheral blood. This suggests there may be a constraint on the sequence diversity due to strong selection for sequences capable of growth in macrophages. In four of the patients, the brain-derived sequences clustered into a separate phylogenetic group from the blood derived sequences, suggesting tissue-specific compartmentalisation of the viruses. However, in two of the six patients, brain-derived sequences clustered together with those derived from the blood, suggesting in these cases that there was no compartmentalisation of the virus, and the possibility of trafficking between the two tissues.

Indeed a number of studies have found that the V3 loop sequence of variants in non-lymphoid tissues are relatively homogeneous (compared to the population found in peripheral blood or lymphoid organs) and show similarity to the macrophage (subtype B) consensus sequence (Power *et al*, 1994; Itescu *et al*, 1994, Ball *et al*, 1994; Li *et al*, 1992).

1.6. The Humoral Immune Response to HIV-1 Infection

1.6.1 The Immune Response to Viral Infections

The immune response to viral infections has three principal components: i) the destruction of virus-infected cells; ii) the neutralising of the infectivity of the virions, and; iii) the production of interferons (White and Fenner, 1994). Cytotoxic T-lymphocytes are responsible for the destruction of infected cells, thus eliminating virus from its reservoirs in the body. Neutralising antibodies neutralise the infectivity of cell-free virus, preventing infection from being established or disseminated. The relative contribution of these three components to recovery from viral infections is disease-specific. Circulating antibody is believed to play a key role in recovery from a number of viral infections: picornavirus, togavirus, flavivirus and parvovirus (White and Fenner, 1994).

Antigenic Determinants

The regions of protein molecules which interact with specific antibody molecules are

described as epitopes (or antigenic determinants). X-ray crystallography has shown that epitopes contain between 15-22 amino acids. Early studies using polyclonal sera raised to a variety of proteins, led to the distinction of two classes of epitopes:
- **linear** (sequential, continuous) epitopes, which require the correct linear amino acid sequence to be recognised. These epitopes can bind antibody when the protein is denatured, and can be mimicked by short peptide sequences.
- **conformational** (discontinuous) epitopes, which require the native protein folding to bind antibody. The discontinuous epitope comprises residues that are brought together by the folding of the polypeptide chain or the juxtaposition of two separate chains of the native protein. In consequence, discontinuous epitopes cannot be mimicked by short peptides, and conformational epitope will be destroyed when the protein is denatured.

In reality, however, most epitopes are conformational, that is, the epitope will be influenced by amino acids outside the epitope, which determine the conformation of the native protein (Benjamin *et al*, 1984; Laver *et al*, 1990).

Neutralising Antibodies

Antibodies bind to accessible epitopes on the surface protein of the virion. However, only those that bind with high avidity to particular epitopes are capable of neutralising viral infectivity. Neutralising antibodies are measured by their ability to inhibit viral infection *in vitro*, through the inhibition of cell-free viral infection, or the inhibition of cell-to-cell infection. Neutralisation is believed to occur by blocking attachment to the host cell, or at a post-binding stage, such as blocking the fusion of the viral and host-cell membranes, thus preventing release of the nucleocapsid.

1.6.2 Characterising the Neutralising Antibody Response Following HIV-1 Infection

Individuals infected with HIV-1 usually develop humoral responses to the majority of viral proteins (Barin *et al*, 1985, Robey *et al*, 1985). Of these antibodies, only a proportion are capable of neutralising the virus.

The main targets of neutralising antibodies are found in the surface envelope protein, gp120 (Profy *et al*, 1990; Rusche *et al*, 1988; Putney *et al*, 1986). The neutralising antibody response elicited by HIV, is of two types: **type-specific**, and **group-specific**. Type-specific neutralising antibodies are capable of neutralising the strain of virus from which the immunogen derived, but not other divergent strains. This type of neutralising antibody is the form predominantly elicited following the experimental inoculation of animals with preparations of the virus, or with recombinant envelope protein (Weiss *et al*, 1986; Looney *et al*, 1988). Group-specific neutralising antibodies are more broadly reactive, in that they are capable of neutralising a wider range of HIV-1 strains (Weiss *et al*, 1985a; Steimer *et al*, 1991; Kang *et al*, 1991; Berkower *et al*, 1989).

Both type-specific and group-specific neutralising antibody responses are found during natural infection. In general, in early infection the neutralising antibody response is type-specific, but later in infection, a group-specific antibody response occurs, although at lower titre (McKnight *et al*, 1992).

1.6.3 The Principal Neutralising Determinant (PND) of HIV-1

Experimental inoculation of animals with preparations of recombinant gp120, or gp120 -derived polypeptides, results in the production of type-specific antisera, capable of neutralising cell-free and cell-to-cell viral infection (Rusche *et al*, 1988). A single dominant epitope was found to elicit these neutralising antibodies, mapping to a portion of the envelope gene within the third hypervariable (V3) region (Rusche *et al*, 1988; Javaherian *et al*, 1989). Consequently, this region has been named the principal neutralising determinant (PND). This region is predicted to form a hairpin loop structure, which is maintained by a conserved central glycine-proline-glycine (GPG) motif and two flanking disulphide bonded cysteine residues (figures 1.4A and 1.5, above).

Fine-structure mapping with synthetic peptides representing regions of the V3 loop has revealed that the antibody binding site is 8 amino acids at the tip, or crown, of the V3 loop around the conserved consensus GPG motif. Type-specificity in part, can

be explained by the amino acid variation present in the V3 loop between different strains of HIV-1, which may alter antibody binding. Similarly, the presence of some cross-reacting neutralising antibodies, directed at the V3 loop, can be explained in part by sequence conservation at the crown of the loop (Javaherian *et al*, 1990; LaRosa *et al*, 1990).

1.6.4 The Targets of Neutralising Antibodies Elicited During Natural Infection with HIV-1

The first detectable neutralising antibody response after primary HIV-1 infection in humans can occur early, within two to six weeks of primary infection, shortly after seroconversion (Albert *et al*, 1990). In other patients, the development is slower, and may occur between 10 to 45 weeks post-infection. (Arendrup *et al*, 1992; Ariyosha *et al*, 1992; Moore *et al*, 1994b).

The neutralising activity of sera taken early after infection appears to be of a type-specific form. It has a high neutralising activity to contemporaneous autologous virus isolates and to a narrow range of laboratory isolates (Albert *et al*, 1990). Competition studies using synthetic peptides of the V3 loop, have shown that the majority of neutralising activity in early sera is directed to the V3 loop (Zwart *et al*, 1992; Bolognesi, 1993; Profy *et al*, 1990; Chamat *et al*, 1992).

A broader (group-specific) neutralising antibody response can be found later in infection (Weiss *et al*, 1986). This broader neutralising antibody response appears to be due to the development of a new class of antibodies directed at epitopes other than the V3 loop, which are more conserved between different isolates. For example, in the case of an accidental infection of a laboratory worker with the isolate HIV-1_{IIIB}, a broad neutralising antibody response was found in the sera 23 months after infection. Sera taken at this time could inhibit the replication of HIV-1_{IIIB} and the divergent isolate HIV-1_{RF}. The neutralising activity could not be blocked by HIV-1_{IIIB}-specific V3 peptides, but these antibodies were found to block binding of gp120 to soluble CD4 (sCD4). A similar pattern of development of neutralising antibodies was seen to occur in the experimental infection of chimpanzees with HIV-1_{IIIB} (Goudsmit

et al, 1988). Initially, high titre antibodies emerge, which are capable of neutralising the autologous strain, and can be blocked by HIV-1_{IIIB}-specific V3 synthetic peptides. After 20 months the neutralising activity broadened to neutralise divergent strains of HIV-1. Again, this set of broader-reacting antibodies were found to block the binding of gp120 to sCD4.

Studies in which HIV-1 positive sera are fractionated by affinity chromatography, using either native recombinant gp120 or denatured gp120, or sCD4, have confirmed that the CD4-binding site is a major target of group-specific antibodies elicited during the course of natural infection and represents a conformationally-dependent epitope (Kang *et al*, 1991; Steimer *et al*, 1991; Berkower *et al*, 1989; Posner *et al*, 1991; Thali *et al*, 1993; Chamat *et al*, 1992).

A neutralising epitope in the V2 loop of gp120 has been defined monoclonal antibodies (Fung *et al*, 1992), and similarly for an epitope in the C2 region (Ho *et al*, 1988). There is a single neutralising epitope in gp41 (Muster *et al*, 1994). $p17^{gag}$ - the matrix protein elicits neutralising antibodies, and may be present on the virion surface (Papsidero *et al*, 1989). However, p17 has amino acid similarity to thymosin a, a human cellular protein, which may be the target of the neutralising antibodies if it is present on the virion surface.

1.6.5 The Role of Neutralising Antibodies in Control of Disease

Whilst there is clear evidence that HIV-1 elicits a strong neutralising antibody response, there are few reports that show the role of neutralising antibodies in controlling HIV-1 infection, or influencing disease progression.

The rapid drop in plasma viraemia seen shortly after primary infection suggests a role for the immune system in counteracting viral replication at this stage of disease. The earliest detectable anti-viral immune response, is a cytotoxic T-lymphocyte response (CTL), with neutralising antibody responses occurring many weeks after the clearance of plasma viraemia (Koup *et al*, 1994; Moore *et al*, 1994b). Similarly, in SIV-infected macaques, a CD8+ T-cell mediated inhibition of SIVmac replication in autologous CD4+ T-lymphocytes was shown to precede the appearance of a neutralising antibody

response, and coincide with the clearance of plasma viraemia (Niu *et al*, 1993).

Despite the initial control of viral replication, high levels of virus are seen in the peripheral blood during the asymptomatic stages of infection, and high viral loads detected in the lymphoid organs (Donaldson *et al*, 1994a; Embretson *et al*, 1993).

A number of studies have addressed the question of whether there is a correlation between the rate of disease progression and the levels of neutralising antibodies. For example, Fenyo, 1994, examined a cohort of HIV-1 infected patients, who could be divided into non-, or slow-progressors and fast-progressors, according to the rate of CD4+ lymphocyte decline over time. Virus isolates from non-, or slow-progressors, in 5 out of 10 cases, could be neutralised by autologous sera. Follow-up sera, taken 3 to 56 months after virus isolation, could in the majority of cases (13 out of 14 sera), neutralise the autologous virus isolates. In contrast, autologous neutralising antibodies were rarely found in the fast-progression group. These studies show that in fast-progressors, the ability to produce neutralising antibodies is severely impaired. They do not however, demonstrate a protective role for neutralising antibodies. The slow rate of progression in patients showing the presence of neutralising antibodies, may be due to another cause, such as a strong CTL response, which was not investigated in this study.

In another study, Lu *et al*, 1993, failed to show a correlation between the presence of neutralising antibodies and the rate of disease progression. In this case, approximately half of patients showing a stable CD4+ lymphocyte count (non-progressors) had high neutralising antibodies, but in the remaining half neutralising antibodies could not be detected. The group of patients showing fast-progression showed undetectable levels of neutralising antibodies, but high levels of plasma viraemia and infected PBMCs. Perhaps surprisingly, within the group of non-progressors, those patients showing high neutralising antibody levels showed the highest levels of virus both in PBMCs and plasma. In contrast, those patients with undetectable neutralising antibody responses had low virus levels. This led the authors to hypothesise that a threshold level of virus may be needed to elicit a neutralising antibody response, but casts doubt on the role of antibodies in controlling

viraemia.

There has been some suggestion that neutralising antibodies provide protection against transmission of HIV-1 from mother to child. In a study of Scarlatti *et al*, 1993, it was found that mothers who did not transmit to their child were more frequently found to have neutralising antibodies to autologous virus, in comparison to mothers who did transmit. Indeed, in respect to this, it has been suggested that a single broadly-specific neutralising antibody, or polyclonal sera, may be used for passive immune therapy to interrupt maternal-foetal transmission. A similar passive immune therapy approach has proved successful in preventing infection in chimpanzees (Emini *et al*, 1991), where the administration of V3-specific antibodies alone was shown to prevent infection of chimpanzees challenged with high doses of cell-free virus.

1.6.6 The Generation of Antibody Escape Mutants

One mechanism by which HIV-1 may evade the immune system is to continually produce novel antigenic variants, which are no longer recognised by the current antibody response. Such antibody-escape mutants have been generated in the laboratory, as will be described below. There is also some evidence that escape mutants arise during the course of natural infection, and in experimental infections of monkeys.

The Generation of Neutralising Antibody Escape Mutants In Vitro

Incubating virus isolates, or biological clones, with small amounts of neutralising antibodies will select for variants that are more resistant to neutralisation. The amino acid changes conferring the increased resistance can then be ascertained by sequencing.

One of the first reports of the generation of antibody escape mutants was by McKeating *et al*, 1989, who cultured virus derived from the molecular clone HIV-1_{NY5}, in the presence of monoclonal antibodies (MAbs) directed to an epitope

bridging the GPG-motif of the V3 loop (figure 1.5, above). A panel of viral mutants was generated, each showing a reduced binding affinity for the MAbs. Four of the variants were found to have an amino acid change in the V3 loop itself (arginine (R) to glycine (G) at position 8), with the corresponding V3 loop peptides unable to bind the selecting antibodies. Other mutants showing reduced sensitivity to neutralisation, however, did not contain amino acid changes in the V3 loop. This result shows that neutralisation by antibodies directed at the V3 loop can be influenced by amino acid changes both within and outside the loop.

In a similar study, using plaque-cloned virus and selecting with a MAb directed at the tip of V3 loop, resistant viruses were generated each of which had a mutation at the tip of the loop, (of GPGR to GGGR), which abolished MAb binding to the corresponding peptide (Masuda *et al*, 1990). Significantly, in the acute-replicating conditions which were required to generate these resistant mutants, there was a large number of amino acid changes in clones which were not incubated in the presence of neutralising antibody.

Resistant viruses can also be generated *in vitro* in response to selection by broadly neutralising antisera. For example Robert-Guroff *et al*, 1986, selected a mutant derivative of the molecular clone HXB2 (Fisher *et al*, 1985), by incubation with a broadly neutralising antisera derived from pooling the sera of HIV-1 infected individuals. The mutation conferring resistance was mapped to the gp41 protein, with a threonine (T) to adenine (A) substitution in the external domain (residues 582 of HXB2) (Klasse *et al*, 1993). It was subsequently shown that this resistant clone was resistant to neutralising antibodies directed at the CD4-binding site (Thali *et al*, 1994). However, only minor differences in the binding of these antibodies to wild type and mutant viruses was observed, showing that the antigenic structure of gp120 can be subtly affected by amino acid changes in gp41, with important consequences for neutralisation. Similarly, Back *et al*, 1993, showed that neutralising resistant variants, arising after the experimental inoculation of chimpanzees with the isolate HIV-1_{IIIB}, have mutations in the extracellular portion of gp41. By generating recombinant viruses, it was shown that these mutations conferred resistance to HIV-1_{IIIB} V3 MAbs.

McKeating *et al*, 1993, showed that escape mutants selected by incubation with HIV-1-positive human serum, developed resistance to CD4-binding site MAbs. The amino acid change conferring resistance was located in the conserved region C4. The mutant envelope protein had reduced affinity to CD4-binding site MAbs, which is in contrast to the result of Thali *et al*, 1994, described above. Clearly, CD4-resistant viruses can arise by two mechanisms: one which directly reduces the affinity of the neutralising antibodies for gp120, and the other which does not affect binding but may block the steps after binding.

The V2 region of gp120 can also evolve *in vitro* under the selection of neutralising antibodies (Yoshiyama *et al*, 1994). In this study, the primary isolate HIV-1_{RF} was incubated with a neutralising MAb directed at a conformational epitope in V2 loop. Three neutralising escape mutants were generated, each showing reduced affinity for the neutralising antibody. A mutation in the V2 loop of Y to histidine (H) (in the motif KEYALFY), was demonstrated as responsible for conferring the resistant phenotype.

However, even in the absence of selection, there can be antigenic variation in gp120 following prolonged culturing. Moore *et al*, 1993c, showed by examination of several molecular clones derived from the laboratory isolate HIV-1_{IIIB}, that each showed a distinct pattern of neutralisation to a panel of neutralising MAbs, including those directed at the V3 loop, the V2 loop and the CD4-binding site. A similar result was described by Domingo *et al*, 1993, who passaged plaque-purified virus in the absence of antibody, and showed that virus progeny displayed considerable sequence and antigenic variation.

Clearly, the high mutation rate of HIV-1 can result in the generation of antigenically distinct variants *in vitro*, even in the absence of immune selection (Domingo *et al*, 1993).

The Development of Neutralising-Antibody Escape Mutants In Vivo

Antigenic variation in non-primate, animal lentiviral infections has been well documented (Lutley *et al*, 1983; Thormar *et al*, 1983), and is frequently quoted as an

important mechanism by which such viruses persist in their host during a chronic infection in the face of a strong anti-viral immune response (Leigh Brown and Holmes, 1994). However, even in these examples, which are often used as paradigms of immunological escape, the significance of antigenic variants in allowing persistent infection, and the role of these variants in pathogenesis has not been unequivocally demonstrated.

For example, in the experimental infection of sheep with visna virus, isolates resistant to autologous sera are seen to emerge, but only at low frequency, and do not replace the infecting strain (Thormar *et al*, 1983; Lutley *et al*, 1983). Virus can persist and cause tissue damage, despite the presence of high titres of neutralising antibodies.

Burns *et al*, 1993, have shown that following the experimental infection of Rhesus macaques with a pathogenic molecular clone of SIV (SIV_{mac239}) *env* gene mutants arise that are resistant to serum neutralisation.

The development of neutralising antibodies following the experimental infection of chimpanzees with HIV-1 has been documented (Nara *et al*, 1990). Virus isolated from chimpanzees vaccinated with gp120 and subsequently challenged with HIV-1_{IIIB}, was found to be more resistant to neutralisation by autologous serum, than the virus isolated from naive animals. This suggests that the immune system, stimulated following vaccination, has selected for more resistant viruses in the vaccinated animals. In both naive and immunised animals, sequential isolates were more resistant to neutralisation by HIV-1_{IIIB}-V3-specific MAbs than the parental stock. Resistance to V3-specific MAbs was conferred by mutations outside the V3 loop. With one isolate, the resistance mutation mapped to gp41 (Back *et al*, 1993).

In the case of the accidental infection of a laboratory worker with the isolate HIV-1_{IIIB}, variants detected one year after infection were found to be resistant to a panel of MAbs directed at the V3 loop (Di Marzo Varonese *et al*, 1993). However, these variants were detected after the development of the variant-specific neutralising antibody response, and persisted in all subsequent isolates (Reitz *et al*, 1994). This suggests that these variants were not selected by neutralising antibodies.

Arendrup *et al*, 1992, and Albert *et al*, 1990, have examined the development of the neutralising antibody response in patients following primary infection. In the first of these, neutralising antibodies to isolates taken one week after seroconversion could be demonstrated only after between 13 to 45 weeks post-seroconversion. Virus isolates with reduced sensitivity to autologous sera taken at or before the time of isolation could be demonstrated at between 45 to 117 weeks. The patients subsequently developed neutralising antibodies to the 'escaped' virus isolates, but the titres against the later isolates were generally low compared to the titres against the first isolate.

In the study of Albert *et al*, 1990, a low-titre neutralising antibody response could be detected, much earlier than the study of Arendrup *et al*, 1992, within 3 to 4 weeks after primary infection. Virus variants with reduced sensitivity to autologous sera was observed. However, in this case, the patients did not develop antibodies capable of neutralising the new variants. In a subsequent study, only a minority of the resistant isolates were found to have amino acid changes in the V3 loop (Wahlberg *et al*, 1991).

1.7 Vaccine Development

Much effort has been directed to the development of vaccines against HIV-1 and HIV-2, a method which has proved effective in controlling the development of disease in a number of other viral infections, such as polio, smallpox, rabies, influenza, hepatitis B and measles.

Traditionally, there have been two major strategies of producing viral vaccines: live, attenuated virus (such as those for polio, measles, mumps), and chemically inactivated virus (such as those for polio, influenza and rabies). With the advent of recombinant DNA techniques, subunit vaccines are being developed to a number of pathogens, and have proved successful in vaccinating against the hepatitis B virus.

With these as precedents, there are currently three main types of vaccine being

produced for HIV and the animal models including SIV infection of rhesus macaques:

- subunit vaccines, using genetically engineered subfragments of the gp120 protein, or synthetic polypeptide epitopes of gp120.
- whole inactivated preparations of HIV-1, or SIVmac
- live, attenuated strains of SIVmac (and HIV-1).

There are several obstacles to HIV vaccine production. In particular, the nature of persistent infection, where HIV (and other lentiviruses) induces a chronic debilitating disease following long-term infection, despite an initially strong host immune response. It is therefore likely that an HIV vaccine will have to produce sterilising immunity against the initial infection, as once the virus is established, there is no epidemiological evidence that the immune response can influence the course of infection or the disease process. However, the very nature of retroviral infection , where a long-lived reservoir of infected cells can be generated by the integration of the provirus into the host genome, and the spread of virus by cell-to-cell contact, undisturbed by the immune defence, makes sterilising immunity following primary infection difficult to envisage (Sabin, 1992).

A second barrier to HIV vaccine production is the antigenic variability displayed by HIV strains. There are a large number of strains of HIV that are minimally cross-neutralising. In consequence, the immune response to one strain may not be effective against the multitude of field isolates which may be encountered by the vaccinee. To date, it has not been possible to quantify the number of HIV strains that would be needed to provide broad protection against infection.

Subunit vaccination of chimpanzees and macaque monkeys has been demonstrated to limited success (reviewed by Bolognesi, 1993). In experiments with chimpanzees in which protection has been demonstrated, protection was found to correlate with the induction of neutralising antibodies directed to the V3 region of gp120. However, in experiments with macaque monkeys, although protection was achieved, there was no correlation between protective immunity and protection.

Protection has also been achieved in macaques following intravenous infection with inactivated SIVmac that had been grown in human T-cells. However, protective immunity was found to be linked to the immune response to human cellular antigens, specifically, HLA-DR and β-micoglobulin, expressed upon the virion envelope, following their growth in human cells (Stott *et al*, 1991). In subsequent experiments, protection was not established with inactivated SIVmac that had been grown in macaque cell cultures.

The most promising vaccine, produced to date, is that of an attenuated strain of SIVmac. Vaccination with the attenuated strain of SIVmac was shown to confer protection in rhesus macaques to high-dose challenge with a pathogenic SIVmac strain (Daniel *et al*, 1992).

The development of a live attenuated SIVmac vaccine came about following experiments carried out to determine the role of infection *in vivo* of the SIVmac *nef* gene. Kestler *et al*, 1991, showed that six rhesus macaques infected with cloned SIV_{mac239} containing a deletion in the *nef* gene were able to maintain an extremely low viral burden and normal CD4+ cell counts, despite the death of eleven out of twelve macaques infected with the wild-type SIV_{mac239}. Four of the monkeys infected with the *nef*-deleted SIV were challenged with wild-type pathogenic SIV_{mac239}, using 10 cell-free infectious doses. Following challenge, these monkeys showed no signs of viral infection with the wild-type challenge virus: no plasma viraemia could be detected, and no wild-type viral sequences, representing the challenge virus, were detected by PCR, despite the amplification of *nef*-deleted sequences. It was concluded that the infection of these monkeys with attenuated, *nef*-deleted SIV_{mac239} rendered them resistant to further virus challenge.

The feasibility of developing a live, attenuated HIV vaccine, with its attendant ethical and safety issues, was discussed by a World Health Organisation working group. The main conclusions of this working group were, that the development of live attenuated vaccine should go ahead in parallel with other vaccine strategies. The safety issues discussed were the stability of such viruses, whether, with the high mutability of primate lentiviruses, they could mutate back to a virulent form, or whether infection

with such viruses will cause delayed immunosuppression, neuropathogenesis, oncogenesis, or other adverse effects. In terms of the efficacy of such vaccines, the breadth of protection against heterotypic strains must be determined, as well as the efficacy against challenge by various routes, dosages and cell-associated virus. In terms of protection, the mechanism and correlates of protection must be determined.

The laboratory of Desrosiers, from which these experiments were first reported, have been constructing multiple-deletion mutants of both SIVmac and HIV-1, in order to render these candidate attenuated viruses, less susceptible to back mutation to the wild-type pathogenic form. To date, it has been shown that HIV-1 viruses carrying multiple deletions in the LTR, *nef*, *vpr* and *vpu* genes, were capable of growth in human CD4+ T cells (Desrosiers, 1992; Gibbs *et al*, 1994).

In 1994, Desrosiers expressed the belief that live, attenuated viruses will be the only effectual vaccine against HIV infection. However, recent work has questioned the safety of attenuated viruses as vaccines. In 1995, Baba *et al*, demonstrated that an attenuated strain of SIVmac is highly pathogenic to neonatal macaques following mucosal infection. Despite containing deletions in *nef*, *vpr*, and the negative regulatory element, the construct caused high levels of viraemia, haemolytic anaemia, thrombocytopenia, and CD4 T-cell depletion, in neonatal macaques.

Despite multiple deletions, and their lack of pathogenicity in adult macaques, these constructs have retained their pathogenicity in infected neonates. As such, these constructs are unsuitable as candidate live, attenuated virus vaccines against AIDS.

1.8 The Work Presented in this Thesis

The work described in this thesis is an experimental investigation of the hypothesis that the *env* gene sequences that emerge during the course of natural infection with HIV-1, represent immunological escape mutants. The work specifically addresses the role of the neutralising-antibody response in selecting *env* gene variants.

In order to demonstrate that the *env* gene sequences of HIV-1 evolve in response to selection from neutralising antibodies, it will be necessary to show that high titres of sequence-specific neutralising antibodies arise subsequent to, but not before, the first appearance of a new variant.

The approach I have taken is to construct a panel of recombinant viruses, expressing the variant *env* genes found over the course of natural infection from a single HIV-1-infected patient. Chapter 3 describes the construction of an HIV-1 proviral vector, into which the *env* genes were substituted to form recombinant viruses. Chapter 4 describes the cloning of the major *env* gene sequences found during a seven-year period of infection, and their incorporation into recombinant proviral clones.

Chapter 5 examines the sequences of the hypervariable regions of the cloned *env* genes, and characterises these according to the *env* sequences previously obtained from this patient (Simmonds *et al*, 1991; Holmes *et al*, 1992). Chapter 6 examines the results of the transfection of the recombinant proviral clones into cell culture in order to recover a panel of recombinant viruses, differing only in the *env* gene.

However, as will be found in chapter 6, the majority of these clones did not yield infectious viruses following transfection, and in addition, the clones that did yield infectious viruses, produced only low titre virus stocks, insufficient to carry out neutralisation assays. Chapter 7, therefore, presents a discussion of this approach, and the possible reasons for the failure of many of the *env* genes to support viral replication in this recombinant system.

Chapter 2

Materials and Methods

2.1 Collection and Processing of Patient Peripheral Blood Samples

Patient 82 is a member of the Edinburgh Haemophiliac Cohort, who were infected with HIV-1 following treatment with a batch of contaminated Factor XIII, prepared from Scottish blood donations (Ludlam *et al*, 1986). p82 seroconverted in June 1984, and was asymptomatic and had not received any anti-viral therapy at the times of collection of the blood samples used in this study. Clinical data for this patient are given by Simmonds *et al*, 1991.

Dates of Collection

Five sequential PBMC samples from p82 were used in this study. The samples were collected in August 1987 (sample #108, year 3, 1987), January 1988 (sample #12, year 4, 1988), February 1989 (sample #82, year 5, 1989A), October 1989 (sample #123, year 5, 1989B), and April 1990 (sample #139, year 6, 1990). No PBMC sample was available at the time of seroconversion (year 0, 1984).

Blood samples were processed by staff at the Centre for HIV Research, University of Edinburgh. PBMCs were isolated from heparinised whole-blood by separation over Ficoll, and stored in liquid nitrogen (Simmonds *et al*, 1990a).

Preparation of PBMC DNA

Ampoules of stored PBMCs (not less than 5 X 10^6 cells) were removed from liquid nitrogen and allowed to thaw. The cells were washed in 20ml RPMI medium (Gibco), and resuspended in 400ul of a lysis buffer containing 50mM EDTA, 100mM Tris-Cl pH 7.5, and 50mM NaCl, 100mM proteinase K to 100mM, and 1% (w/v) N-laurylsarcosine. Lysis was allowed to proceed by incubating the reaction at 65^{0}C for 2hr. The DNA was extracted by a standard phenol/chloroform extraction, and precipitated by the addition of 2.5 volumes of 100% ethanol and storage at -20^{0}C for at least 2hr. The DNA was collected by centrifugation, washed with 70% ethanol and dried at 50^{0}C. The DNA was resuspended in 100-200ul of dH_20 and the concentration assessed by spectrophotometry.

2.2. HIV-1 Infectious Molecular Clones

pHXB2-D

pHXB2-D is a full-length proviral clone capable of producing infectious viruses when transfected in cell culture (Fisher *et al*, 1985). pHXB2-D was obtained from the MRC AIDS Reagent Project Repository, National Institutes for Biological Standards and Control, to which it was donated by Dr Robert Gallo of the National Institutes of Health, Bethesda, Maryland, USA.

The proviral clone HXB2 was originally isolated in 1984, as a lambda clone from a library of recombinant bacteriophage constructed from the *Xba*I digested genomic DNA of H9 cells chronically infected with the HIV-1 isolate HIV-1_{IIIB} (Shaw *et al*, 1984). A 12.7kb *Xba*I insert fragment of the lambda clone HXB2, consisting of 9.8kb of full-length proviral sequence and 2.7kb of cellular flanking DNA, was transferred into the plasmid vector pSP62 (Melton *et al*, 1984) to create the molecular clone pHXB2-D (Fisher *et al*, 1985).

The complete nucleotide sequence of HXB2 is available from GenBank, accession number K03455. In addition, a schematic diagram of pHXB2-D is given in figure 3.1, and the restriction site co-ordinates for the proviral sequence (HXB2R) are given in appendix A.

NOTE: Sequence analysis has subsequently revealed that the isolate HIV-1_{IIIB} originated from the HIV-1 infected patient LAI, from which virus was originally isolated at the Pasteur Institute, Paris, France. It appears that virus from the isolate HIV-1_{LAI}, contaminated and outgrew cultures of HIV-1_{IIIB} and HIV-1_{BRU}. In consequence, both isolates have been renamed HIV-1_{LAI} (Wain-Hobson *et al*, 1991). However, for the purposes of this thesis, the isolate from which HXB2 was cloned will be referred to by its original name of HIV-1_{IIIB}.

pHXB2-MCS

pHXB2-MCS is a modified version of pHXB2-D. The details of the construction of pHXB2-MCS are described in chapter 3, section 3.3, of this thesis. In brief, the 12.7kb *Xba*I proviral and cellular flanking sequences of pHXB2-D has been transferred to the plasmid vector pSPTBM20Δ - a derivative of the plasmid vector pSPTBM20 (Boehringer) from which the *Eco*RV to *Sma*I fragment of the polylinker has been deleted. Three unique restriction sites have been engineered into the proviral genome: a *Not*I site at 5645, a *Bst*EII site at 5875 and an *Xba*I site at 7624 (co-ordinates according to HXB2R).

pHXB2-MCS.*Mlu*I

pHXB2-MCS.*Mlu*I is a derivative of pHXB2-MCS, containing an additional cloning site - *Mlu*I - at nucleotide position 7279 in the HXB2R sequence. This clone was modified from pHXB2-MCS by Dr Peter Balfe at The University College and Middlesex School of Medicine, Department of Medical Microbiology, London. Further details of its construction can be found in chapter 3, section 3.3.

$SF2_{MC}$ and $SF162_{MC}$

These are two full-length HIV-1 proviral clones, a gift from Dr Cecelia Cheng-Mayer at the School of Medicine, Cancer Research Institute, University of California, San Fransisco, USA.

$SF2_{MC}$ is derived from the HIV-1 isolate HIV-1_{SF2}, which was obtained from the co-cultivation of PBMCs of a symptomatic HIV-1 infected individual with oral candidiasis, with PBMCs from seronegative donors (Cheng-Mayer *et al*, 1989). $SF162_{MC}$ is derived from the HIV-1 isolate HIV-1_{SF162}, which was obtained by the co-cultivation of cerebrospinal fluid of an HIV-1 infected individual with neurological symptoms (toxoplasmosis) with PBMCs from seronegative donors (Cheng-Mayer *et al*, 1989).

Both full-length clones were originally obtained from libraries of recombinant bacteriophage constructed from the *Eco*RI-partially-digested genomic DNA of the

respective isolates. The full-length versions of the molecular clones $SF2_{MC}$ and $SF162_{MC}$ were divided into two fragments at unique internal *Eco*RI sites, at the midpoint of each genome. The 5' and 3' viral fragments generated after restriction digest with *Eco*RI were subcloned into a pUC19 plasmid vector at the *Eco*RI site (Cheng-Mayer *et al*, 1990).

The complete nucleotide sequence of $SF2_{MC}$ is available from Genbank, accession number K02007. The nucleotide sequence of the 3' half of the $SF162_{MC}$ genome is available from GenBank accession number M38428.

2.3 Plasmid Cloning Vectors

pBluescriptII(KS-)

pBluescriptIIKS(-) (Short *et al*, 1988, Alting-Meese and Short, 1989) is a 2.96kb plasmid cloning vector and was purchased from Stratagene Ltd. This vector contains a ColEI origin for autonomous replication in the *E.coli* host, and an f1 filamentous phage origin of replication, allowing recovery of single stranded copies of the plasmid sequence (- the antisense strand of the *lacZ* gene - encoded by the vector), on infection with helper phage. pBluescriptIIKS(-) has a T7 RNA polymerase promoter flanking the polylinker. The complete sequence of pBluescript is available from GenBank accession number X52329.

pSPTBM20

The plasmid cloning vector pSPTBM20 was purchased from Boehringer Mannheim.

pNBXXΔ*env*

pNBXXΔenv is an *env* gene subclone of pHXB2-MCS. Its construction is described in chapter 3, section 3.6. In brief, the *Not*I to *Xho*I fragment of pHXB2-MCS (nts 5645-8443 HXB2R) was subcloned into pBluescriptIIKS(-) at the *Not*I and *Xho*I sites in the vector polylinker, generating the construct pNBXX. A deletion of 428bp was made in the *env* gene by double digestion with *Stu*I (nt5952 HXB2R) and *Nde*I (nt6380 HXB2R) followed by filling-in the 3' overhangs with Klenow and self-

ligation.

pHXB2-MCSΔ*env*

pHXB2-MCSΔenv is a deletion-derivative of pHXB2-MCS. A 428bp deletion was made within the *env* gene - from *Stu*I (nt5952 HXB2R) to *Nde*I (nt6380 HXB2R). Transfection of pHXB2-MCSΔenv into cell culture results in the production of $p24^{gag}$ antigen, but does not generate infectious viruses (chapter 6, section 6.2.2).

2.4 Bacterial Strains

The *E.coli* strains used in this study are given in table 2.1 below.

Bacterial Stocks

Bacteria were maintained short-term on solid medium (generally, 2xTY agarose plates see appendix B), at 4^0C. For longer-term storage, DMSO stocks were prepared, and could be stored for over a year at -20^0C or -70^0C. These were prepared by adding 1ml of a freshly saturated culture to an equal volume of 7% (v/v) DMSO solution. Fresh colonies were obtained by streaking bacteria from single colonies on existing plates, or by streaking a loop-full from the frozen stock onto a fresh plate and incubating overnight at 37^0C The presence of the F' episome, required for the infection of R408 M13-derived helper phage, was ensured by streaking colonies on 2xTY plates with supplemented with tetracycline.

Storage of Plasmid DNA

Plasmid DNA constructs were stored with their *E.coli* host at -20^0C, and as DNA preparations at -20^0C. Due to the instability of HIV-plasmid constructs in their *E.coli* hosts, and the poor viability of strain DL655, plasmids were also stored as ethanol precipitates at -20^0C.

Table 2.1 *E.coli* Strains

Strain	Genotype	Source	Reference
DL655-F'	*hsd*R, *leu*, *mcr*A, *mcr*B, *pro*, *rec*D 1009, *sbc*C201, *sup*E, *thi*-1, *tsx*, (*dcm*, *dam*) *rec*A::Cm^r [F' *pro*AB, *lac*I^{q}ZΔM15, Tn10(tet^r)] DL655-F' is DL655 transfected with the F' episome of XL1-Blue (Stratagene Ltd).	DL655 was a gift from David Leach at ICMB, University of Edinburgh.	Chalker *et al*, 1988 Bullock *et al*, 1987 (XL1-Blue)
SURE	e14$^-$ (*mcr*A), Δ(*mcr*CB-*hsd*SMR-*mrr*)171, *sbc*C, *rec*B, *rec*J, *umu*C::Tn5(kan^r), *uvr*C, *sup*E44, *lac*, *gyr*A96, *rel*A1, *end*A1 [F' *pro*AB, *lac*I^{q}ZΔM15, Tn10(tet^r)]	Stratagene Ltd	Greener *et al*, 1990
XL1-Blue	*end*A1, *hsd*R17, *sup*E44, *thi*-1, *rec*A1, *gyr*A96, *rel*A1, *lac*, [F'*pro*AB, *lac*I^{q}ZΔM15, Tn10(tet^r)]	Stratagene Ltd	Bullock *et al*, 1987
CJ236	*dut*1, *ung*1, *thi*-1, *rel*A1/pCJ105(Cm^r), [F' *pro*AB, *lac*I^{q}ZΔM15, Tn10(tet^r)]	Invitrogen	Joyce and Grindley, 1984

2.5 DNA Manipulations and Cloning

Small-Scale Preparation of Plasmid DNA

The boiling method of Holmes and Quigley (Homes and Quigley, 1981), was used to prepare small amounts (2-3ug) of plasmid DNA from *E.coli*.

An overnight culture of *E.coli*, carrying the required plasmid, was prepared by inoculating 5ml of 2xTY-broth plus ampicillin with a single bacterial colony, and incubating at 37^{0}C overnight with vigorous aeration. 1.5ml of the culture was centrifuged for 30s to pellet the cells. The supernatant was poured away, and the pellet resuspended in 300ul STET solution (see appendix B) and placed on ice. To this was added 30ul of a freshly-prepared 10mg/ml lysozyme solution (Boehringer). The reaction was incubated on ice for 2min and heated to 100^{0}C for 90s. Immediately after this, the sample was centrifuged at 1500g for 15min. The resultant flocculant pellet of bacterial debris was removed with a sterile toothpick. 0.7 volumes of isopropanol and 0.1 volume of 3M NaAc was added, and the sample stored at -20^{0}C for 20min to precipitate the DNA. The precipitate was collected by centrifugation at 1500g for 10min, washed in 70% ethanol and allowed to air-dry. The DNA was resuspended in 50ul of TE. The concentration was determined by agarose-gel electrophoresis of a 5ul aliquot.

Restriction Digests

Restriction digests were carried out using between 0.2 and 10ug of the appropriate DNA preparation. Between 5 to 10 units of restriction enzyme, were used according to the amount of DNA to be digested, with the restriction enzyme buffer recommended by the manufacturer (Boehringer, Stratagene or Promega). Digests were allowed to proceed for at least 60 min, at the recommended optimum temperature for the enzyme. When two or more enzymes were used to digest the DNA, the broad-specificity buffer KGB was used, at the recommended concentration (McClelland *et al*, 1988). When necessary, for subsequent enzymatic steps, restriction enzymes were inactivated by heating at 65^{0}C for 15 min for heat-sensitive enzymes. For heat-stable, phenol resistant enzymes (such as *Xho*I), the enzyme was removed by purification of the DNA using a 'Magic DNA Clean-up' column (Promega).

Agarose-Gel Electrophoresis

Agarose gels were prepared with standard electrophoresis grade agarose (Sigma), or with low melting point agarose (LMP; Sigma) when high percentage gels (>2.5%) were required and when DNA was to be recovered from the gel. Gels were prepared by dissolving the required amount of agarose in TBE buffer. Once the molten gel had cooled to below 50^{0}C, ethidium bromide was added to 1ug/ml. Ficoll stop buffer was added to the DNA sample prior to electrophoresis. Electrophoresis was carried out at a voltage appropriate to the application. For the optimum separation of fragments within the range of 1 to 3kb, required routinely in this project, the voltage applied was 1.5V/cm. For more rapid separation, voltages up to 5V/cm were used.

Recovery of DNA from Agarose Gels

The majority of the subcloning carried out in this study, involved the isolation and purification of specific restriction fragments. This was achieved by first resolving the restriction fragments by agarose-gel electrophoresis, then recovering the required fragment from the agarose-gel.

To obtain optimum resolution of the DNA fragments (which in the majority of cases were in the range of 0.8 to 2.5 kb), electrophoresis was carried out with 0.5% LMP agarose gels, at between 0.3 and 1 volt/cm. Following electrophoresis, the required bands were excised from the gel using a clean scalpel blade, (or in the case of infectious clones, using a plastic knife), illuminating the DNA under long-wave UV to minimise UV-induced damage to the DNA. The gel-slice was incubated at 65^{0}C until the agarose had completely melted. Purification of the DNA was achieved by treatment with 'Gene-Clean' (Bio 101 inc), according to the manufacturer's instructions, or through the use of a 'Magic DNA clean-up' column (Promega). The recovery of the DNA was assayed by running one tenth of the recovered volume on an agarose gel.

Ligations

The relevant vector and/or insert DNA was prepared by digesting approximately 2-10ug of the plasmid construct with the appropriate restriction enzymes as described above. In general, the cohesive ends produced from the restriction digests of the vector were

compatible with the cohesive end of the insert DNA to be cloned. In situations where it was necessary to fill-in incompatible recessed 3' termini, the Klenow fragment of *E.coli* DNA polymerase I was used. In such cases, 1unit of Klenow (Boehringer) was added to the reaction on completion of the digest, and dNTPs (Boehringer) to 1mM. The reactions were incubated for 30min at 37^{0}C and subsequently terminated by heating to 750C for 10min. The required DNA fragments were then isolated and purified by agarose-gel electrophoresis (as described above).

The relevant vector and insert fragments were ligated in reactions that contained a 2- to 5- molar excess of insert to vector DNA. Due to the low transformation efficiency of *E.coli* strain DL655F', approximately 300ng of vector DNA were required to recover colonies from each ligation reaction. Ligations were carried out in as small a volume as possible (generally between 10-30ul) with 1unit of T4 DNA polymerase (Boehringer), in the presence of 0.5x KGB (McClelland *et al*, 1988; appendix B), ATP to 1mM, or with 1x T4 DNA ligase buffer supplied by the manufacturer (Boehringer, appendix B). Reactions were incubated on ice contained within a glass beaker which was left for 16hr on the open bench, allowing the ice to melt and reach room temperature (15-20^{0}C). The entire ligation reaction was used to transform competent cells of *E.coli* strain DL655F' (table 2.1, above). In this manner, between 20-100% of recovered transformant colonies contained the required recombinant plasmids.

Screening of Recombinant Clones

Two methods were used to screen transformant bacterial colonies for the presence of plasmids successfully incorporating the required insert fragment. The first was to prepare plasmid DNA from overnight *E.coli* cultures initiated from single transformant colonies and analyse the DNA by restriction digest. The second method used was the direct screening of transformant colonies by the polymerase chain reaction (PCR). The PCR-based method described below has taken the cloning of PCR-amplified *env* genes of lengths 1.7kb and 2.5kb, into the plasmid vector pNBXXΔ*env* as the example. The same method was also used to screen for the correct transfer of the cloned *env* genes into the proviral vector pHXB2-MCSΔ*env*.

The Direct Screening of Transformant Colonies by PCR

The screening of transformant colonies for the incorporation of the PCR-amplified DNA fragment takes the advantage of the difference in size between the cloning vector pNBXXΔ*env*, and the recombinant derivatives which have successfully incorporated the PCR fragment. pNBXXΔ*env* contains the entire HXB2 *env* gene minus 428bp at the 5'end of the gene. Recombinant clones incorporating the PCR-amplified *env* gene fragments should have a full-length *env* gene. Transformant colonies will therefore be of two types: there will be those carrying the reconstituted or incompletely digested vector; and secondly those carrying the required PCR-derived *env* gene fragment. These can be distinguished by specifying PCR-primers that amplify across the deletion of pNBXXΔ*env* and using these in a PCR to amplify the DNA directly from the transformant colonies. The primers for the PCR-amplification are '313' and '403' - whose specification is given below in section 2.7. The PCR-product of the reaction will be of length 140bp for pNBXXΔ*env*, and of approximately 560bp for full-length insert-bearing plasmids.

For each screening, PCR buffer was prepared for the appropriate number of replicate reactions, and 20ul was distributed into each PCR tube. Reactions contained 0.33mM each dNTP, 100nM of each primer, 0.025 units/ul *Taq* polymerase (Promega), 50mM KCl, 10mM Tris-Cl (pH 8.0), 1.5mM $MgCl_2$, 0.1% Triton X-100. Single colonies were inoculated firstly into the PCR tube and then immediately into 2ml 2xTY broth plus ampicillin. The cultures were initiated by incubating at 37^0C with vigorous aeration. The PCR was carried out with 25 heat-cycles each consisting of 25s at 94^0C, 35s at 50^0C, and 150s at 68^0C, followed by an extension step of 10min at 68^0C. The PCR products were analysed by agarose-gel electrophoresis on 2.5% agarose gels. For colonies yielding the correct sized PCR-product (560bp), plasmid DNA was extracted from the cultures initiated from the colonies and further characterised by restriction digest.

2.6 Preparation and Transformation of Competent *E.coli* Cells

Preparation of Competent Cells

The preparation of competent cells was carried out by the method of Hanahan (1983). Bacteria are rendered susceptible to the uptake of DNA by treating log-phase cultures to heavy metal ions.

Cells were plated onto 2xTY agarose plates (appendix A) and incubated overnight at 37^{0}C. A single colony was used to inoculate 10ml of 2xTY broth, and the culture incubated for 2hr at 37^{0}C. 5ml were subcultured into 95ml of 2xTY broth and incubated at 37^{0}C until the cells had reached an optical density at wavelength 550nm of between 0.5 and 0.6. The culture was chilled on ice for 5min before aliquoting the contents into pre-chilled corex tubes and chilling on ice for a further 10min. The cells were pelleted by centrifugation at 6,000rpm for 5min in a cold Sorvall centrifuge. The supernatant was discarded and the cells resuspended in 10ml TfbI solution (appendix A). The cells were incubated for 15min on ice and centrifuged at 6,000rpm for 5min. The supernatant was again discarded and the cells resuspended in 1ml of TfbII solution (appendix B). The cells were incubated for 15min on ice before aliquoting 100-300ul into pre-chilled Eppendorf tubes contained on ice. The cells were then snap-frozen in liquid nitrogen and stored at -70^{0}C for subsequent use.

The efficiency of transformation of the competent cells prepared by this method was assessed by transforming an aliquot of cells with a known amount (between 0.1ng and 10ng) of supercoiled plasmid DNA. The efficiency of transformation of competent cells of *E.coli* strain DL655F' used routinely in this project was 10^4/ug supercoiled DNA, whereas the same method produced an efficiency of transformation of 10^6/ug supercoiled DNA from more robust strains such as XL1-Blue (Stratagene).

Transformation of *E.coli* with Plasmid DNA

Aliquots (100 to 300ul) of competent cells were removed from storage at -70^{0}C and thawed on ice. The cells were mixed gently and 100ul aliquoted into pre-chilled 1.5ml Eppendorfs. The appropriate DNA for transformation was chilled on ice and added to the competent cells. The reactions were incubated on ice for 30 to 45min, and heat-

shocked by placing at 45°C for 45s. Immediately after heat-shock the cells were returned to ice for 2min. 0.5ml of 2xTY broth pre-heated at 37°C was added to the cells which were then incubated at 37°C for 1hr. After this time, the cells were plated onto 2xTY plates supplemented with ampicillin and incubated overnight at 37°C.

2.7 The Polymerase Chain Reaction (PCR)

PCR Primers

PCR primers were specified to amplify five regions of the HIV-1 *env* gene:

i) the V1/V2 region (278bp);

ii) the V3 region (289);

iii) a 1.7kb fragment (cassette-1);

iv) a 2.5kb (cassette-2); and

v) a 1.4 kb region (cassette-3)

- with the first four using a nested PCR. The lay out of the primers is shown in figure 2.1. The primers were specified at conserved regions of the HIV-1 genome, using the consensus sequence derived from the Los Alamos Database of HIV-1 subtype B sequences (Myers *et al*, 1991).

V1/V2 Hypervariable Region

outer:

401 5' GAG GAT ATA ATC CAG TTT ATG G (+, 6107)

333 5 GTA CAT TGT ACT GTG GTG ACA (-, 6493)

inner:

402 5' GAT CAA AGC CTA AAG CCA TG (+, 6126)

403 5' CAA TAA TGT ATG GGA AAT TGG (-, 6404)

V3 Hypervariable Region

outer:

332 5' TAC AAT GTA CAC TGG AAT T (+, 6503)

308 5' ATT ACA GTA GAA AAA TTC CCC (-, 6908)

inner:

306 5' TGG CAG TCT AGC AGA AGA AG (+, 6572)

307 5' AAT TTC TGG GTC CCC CTC CTG AGG (-, 6861)

Cassette-1: 1.7kb (BstEII to XbaI)

outer:

365 5' GAT GTT GAT GAT CTG TAG (+, 5856)

366 5' TCC AGG TCA TGT TAT TCC AA (-, 7636)

inner:

313 5' TGG GTC ACC GTC TAT TAT (+, 5891) *Bst*EII

367 5' GGA GTA ATA AAT GTC TAG ATG (-, 7630) *Xba*I

Cassette-2: 2.5kb (BstEII to XhoI)

outer:

365 5' GAT GTT GAT GAT CTG TAG (+, 5856)

638 5' GAT GAA CAC TAA CGA GGT AC (-, 8462)

inner:

313 5' TGG GTC ACC GTC TAT TAT (+, 5891) *Bst*EII

g1999 5' GTC TCG AGA TAC TGC TCC (-, 8432) *Xho*I

Cassette-3: 1.4kb (BstEII to MluI)

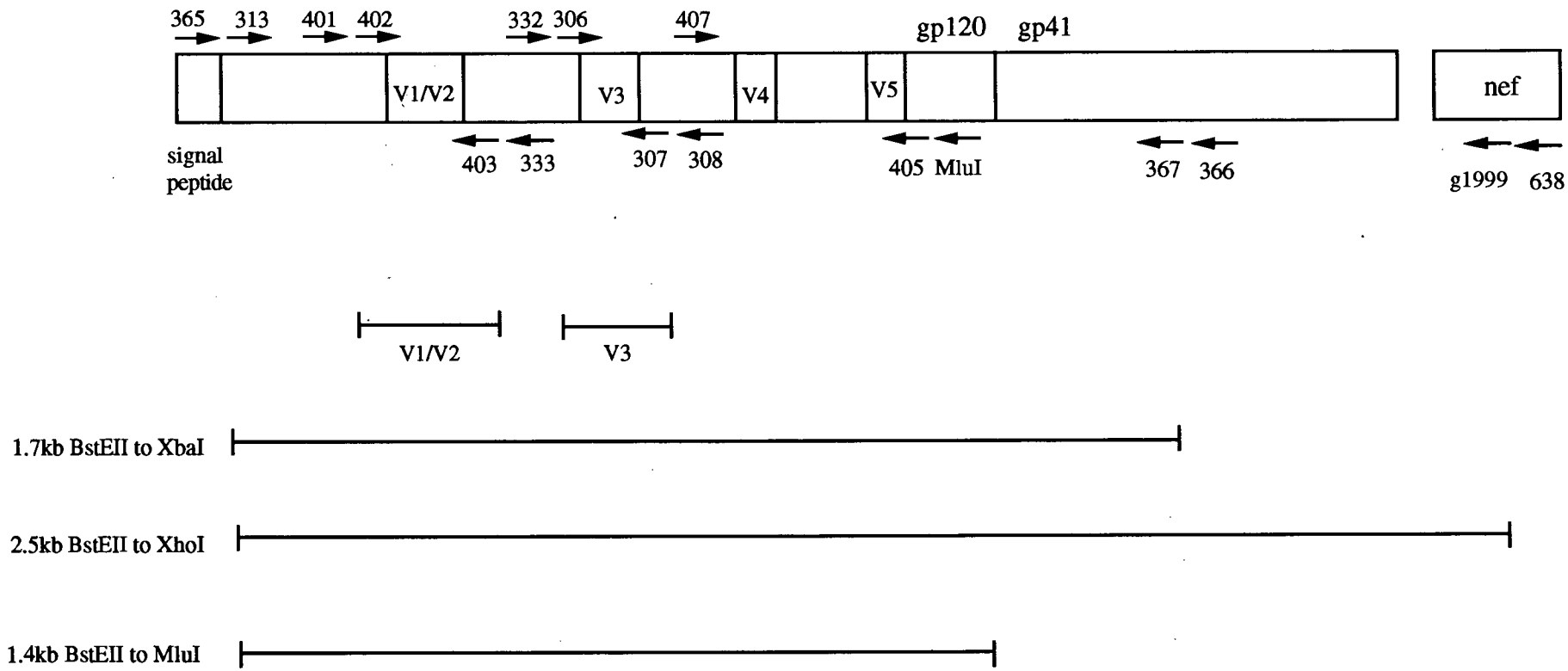
313 5' TGG GTC ACC GTC TAT TAT (+, 5891) *Bst*EII

*Mlu*I 5'TGC ACC ACG CGT CTC TTT GCC (-, 7270) *Mlu*I

(The restriction sites located within the primers are underlined)

Limiting-Dilution Nested PCR

The limiting-dilution nested PCR was carried out according to Simmonds *et al*, 1990a. A series of dilutions of the PBMC DNA were made (usually from 1:10 to 1:1000) and these were amplified in replicate in a nested PCR with primers specified to amplify the V1/V2 hypervariable region of the *env* gene (as given above). The amplifications were repeated with different PBMC DNA concentrations until the end-point dilution was


365
313
401
402
332
306
407
gp120
gp41
V1/V2
V3
V4
V5
nef
signal peptide
403
333
307
308
405
MluI
367
366
g1999
638
V1/V2
V3
1.7kb BstEII to XbaI
2.5kb BstEII to XhoI
1.4kb BstEII to MluI

reached where less than 20% of the PCR reactions yield amplified DNA.

The primary and secondary reactions contained 0.33mM each dNTP, 100nM of each of the appropriate primer pair, 0.025 units/ul *Taq* polymerase (Promega), 50mM KCl, 10mM Tris-Cl (pH8.0), 1.5mM $MgCl_2$, 0.1% Triton X-100. For each primary reaction 1ul of diluted PBMC DNA was amplified in a 20ul volume. For the secondary reaction 1ul from the primary reaction was transferred into a 20ul volume. Each set of amplifications was set up in parallel with 5 negative controls in which the DNA was omitted. For each round of amplification, 25 heat-cycles were used, each consisting of 25s at 94°C, 35s at 50°C, and 150s at 68°C, followed by an extension step of 10min at 68°C. The products of the secondary reaction were analysed by agarose gel electrophoresis.

PCR Amplification with *Pfu*-Polymerase

To amplify the cassettes from single proviral molecules directly from PBMC DNA, a nested PCR with 1ul of PBMC DNA at the end-point dilution calculated from the V1/V2 PCR was employed. The primary and secondary rounds of amplification were carried out with 100uM each dNTP, 0.025units/ul *Pfu* polymerase (Stratagene), 20mM Tris-Cl (pH 8.8), 10mM KCl, 6mM (NH4)SO4, 1.5mM MgCl2 and 0.1% Triton X-100 and 100nM each of the appropriate primer pair. Amplifications were carried out by 25 heat-cycles of 94°C for 35s, followed by 45°C for 35s and 72°C for 4min. After the 25 cycles the reaction were held at 72°C for 10mins to allow completion of strand synthesis.

The primary round of amplification was carried out in a 20ul volume. On completion, 1ul was transferred to the secondary round amplification. The initial secondary round amplification - for the purpose of screening for the presence of amplified DNA - was carried out in a 20ul reaction. The PCR products were analysed by subjecting the entire 20ul volume to agarose gel electrophoresis on 0.8% agarose gels. Reactions containing amplified DNA of the appropriate size were re-amplified from 1ul of the corresponding primary reaction in a 100ul volume. Following the secondary PCR, 10ul was analysed by agarose gel electrophoresis and the remaining 90ul cloned as described below.

When sequences were amplified from cloned material, a single round of PCR with the appropriate inner primers was used, in a 100ul volume containing 10ng of the appropriate cloned DNA.

Cloning of PCR Products

Immediately following the completion of the PCR cycles, the reactions were placed on ice to minimise the exonuclease activity present in *Pfu* polymerase. The paraffin oil was removed by the addition of 400ul chloroform, and extracting the aqueous layer. Optimum cloning efficiencies were obtained by purifying the DNA through a simple ethanol precipitation. This was achieved by adding 0.1 volume of 3M NaAc pH 4.5, and 2.5 volumes of 100% ethanol and, to facilitate precipitation, 1ug glycogen was added to the reaction (Tracy, 1981) and the sample stored at -20^{0}C for at least 20min. The precipitate was collected by centrifugation at 1500g for 10min, washed twice with 70% ethanol and allowed to air dry. The DNA was resuspended in a 10ul volume with 5 units of the appropriate restriction enzyme (*Xho*I, *Xba*I or *Mlu*I) and KGB restriction enzyme buffer to a 1X concentration (McClelland *et al*, 1988). Incubation of the sample was carried out at 37^{0}C for 2hr. After this time, the reaction was pulse-centrifuged to collect any condensation on the sides of the tube, and 5units of *Bst*EII was added, the sample overlaid with paraffin oil and incubated at 65^{0}C for 1hr.

Approximately half of the reaction was used to ligate into the plasmid vector pNBXXΔ*env*. The entire ligation reaction was used to transform competent cells of *E.coli* strain DL655F'(as described above).

2.8 Oligonucleotide-Directed Mutagenesis

Oligonucleotide-directed mutagenesis was carried out by an adaptation of the original method of Kunkel, 1985. An oligonucleotide is specified to contain the desired nucleotide sequence changes. These are incorporated into a cloned DNA sequence by priming the *in vitro* synthesis from a single-stranded form of the cloned sequence with the oligonucleotide. The synthesised strand containing the specified changes is

recovered by transformation of *E.coli*. The efficiency of recovery of the mutant over the wild-type strand can be enhanced by selecting against the parent strand. With the method of Kunkel, selection is achieved by preparing the template construct from the *E.coli* strain CJ236 (Joyce and Grindley, 1984), (*dut*, *ung*, table 2.1). This strain incorporates a small number of uracil residues in place of thymine. *dut* mutants lack the enzyme dUTPase and contain raised levels of dUTP, which competes with dTTP for incorporation into DNA; *ung* mutants lack uracil glycosylase which would normally remove uracil from DNA. Following transformation of uracil-containing DNA into a wild type *E.coli* host, the uracil residues will be removed by the action of N-uracil glycosylase, leaving the DNA susceptible to nuclease degradation, and consequently rendered biologically inactive. The mutant strand, lacking uracil should have greater survival in the wild type host.

Preparation of a Uracil-Containing Single-Stranded DNA Template

E.coli strain CJ236 (table 2.1) was transformed with 20ng of the appropriate plasmid construct. A single transformant colony was inoculated into 5ml 2xTY-broth plus ampicillin, supplemented with 0.25ug/ml uridine (Boehringer), and incubated at 37°C with vigorous shaking for 60min to allow expression of the F' episome contained on the plasmid. 50ul of R408 helper phage (Russel *et al*, 1986) at 10^{11}pfu/ml (Stratagene) was added and incubation continued overnight. 1.5ml aliqouts of the culture were centrifuged for 10min at 4°C to pellet the cells. 200ul of a 15% PEG/2.5M NaCl solution was added to 1ml of the supernatant and the mixture left at room temperature for 20min, to allow the phage particles to precipitate. The phage were pelleted by centrifugation at 1500g for 10min and resuspended in 100ul TE. The DNA was extracted by a standard phenol/chloroform extraction followed by ethanol precipitation, and resuspended in 20ul TE. The yield was determined by agarose-gel electrophoresis of a 5ul aliquot.

In Vitro Synthesis of Mutant Strand

200ng of uracil-containing, single-stranded DNA template was annealed with 20ng of the appropriate phosphorylated primer (see below), in a 10ul volume containing 10mM Tris-Cl (pH 7.5), 2mM $MgCl_2$ and 50mM NaCl. Annealing was carried out by heating

the reaction to 75°C followed by gradual cooling to room temperature, over a period of 1 hour. Synthesis of the mutant strand was carried out by the addition of 1ul of a 10X synthesis buffer (4nM dNTPs, 10mM ATP, 200mM Tris-Cl pH 7.5), 40mM $MgCl_2$ and 150 mM DTT, 5 units Sequenase (USB) and 1 unit T4 DNA ligase (Boehringer). The reaction was incubated on ice for 5min, followed by 5 min at 25°C and 37°C for 60min. The DNA was ethanol precipitated after addition of 20ng glycogen (Boehringer) to facilitate DNA precipitation. The DNA was collected by centrifugation, washed in 70% ethanol and allowed to air dry. A further ligation step was carried out by resuspending the DNA in 20ul of 0.5x KGB (appendix A) with 1 unit T4 DNA ligase (Boehringer), ATP to 1mM and incubating for 60min at room temperature.

Recovery of Mutant Clones

One quarter of the ligation reaction was used to transform competent cells of *E.coli* strain SURE *ung*+ (Stratagene) (table 2.1). Colonies were screened for mutant clones by digestion of miniprep DNA, prepared from single transformant colonies.

Primers for Oligonucleotide Mutagenesis

The following primers were specified for oligonucleotide mutagenesis:

- *Not*I: 5' GTA GCA TTA GCG GCC GCA ATA ATA ATA GC 3' (5645 HXB2R)
- *Bst*EII: 5' TGG GTC ACC GTC TAT TAT G 3' (5891 HXB2R)
- *Xba*I: 5' GAA GTA ATA AAT GTC TAG ATG 3' (7630 HXB2R)

The nucleotide sequences were specified to be identical to the sense strand of the HXB2 sequence, except for the appropriate substitutions to generate the three respective restriction sites.

Phosphorylation of Oligonucleotide Primers

Primers synthesised by Oswell contain neither a 3'or 5' phosphate group. In order to improve the efficiency of the ligation of the newly synthesised strand produced during *in vitro* mutagenesis, the oligonucleotide primers were first phosphorylated at the 5' end.

15 pmoles primer were phoshphorylated in a 20ul volume, by incubating with 2 units of T4 polynucleotide kinase (Boehringer) in the presence of 1mM ATP, and T4 polynucleotide kinase buffer (Boehringer) to the final, recommended, concentration of 50 mM Tris.Cl pH8.2, 10mM $MgCl_2$, 0.1 mM EDTA, 5 mM DTT and 0.1 mM spermidine, at 37°C for 30min. Reactions were terminated by heating to 65°C for 10 min.

2.9 DNA Sequencing

2.9.1 Manual Sequencing

DNA sequencing was carried out using the dideoxy-chain termination method of Sanger *et al*, (Sanger *et al*, 1977). The reaction conditions followed those recommended by the Sequenase version 2.0 sequencing protocols (USB), with the addition of 10% DMSO to the annealing and termination reactions according to Winship, 1989.

Primers for Manual Sequencing

For the position of the primers along the *env* gene, see figure 2.1

Hypervariable Regions V1 and V2

402 5' GAT CAA AGC CTA AAG CCA TG (+, 6126)

403 5' CAA TAA TGT ATG GGA AAT TGG (-, 6404)

Hypervaraible Region V3

306 5' TGG CAG TCT AGC AGA AGA AG (+, 6572)

307 5' AAT TTC TGG GTC CCC CTC CTG AGG (-, 6861)

Hypervaraible Region V4 and Conserved Region C4

407 5' GGG GAA TTT TTC TAC TGT AAT (+, 6928)

Hypervariable Region V5

405 5' CTT CTC CAA TTG TCC CTC ATA (-, 7192)

5' End of gp41

367 5' GGA GTA ATA AAT GTC TAG ATG (-, 7630)

Denaturation For each reaction approximately 1-3ug of plasmid DNA was denatured by treatment with alkali, by incubating in a 50ul volume containing 0.2M NaOH and 0.2mM EDTA, for 20min at 37°C. The reaction was neutralised by the addition of 0.1

volume of 3M NaAc pH 4.3. The DNA was precipitated by adding 2.5 volumes 100% ethanol, and storing at -20^{0}C for at least 20min. The DNA was collected by centrifugation for 10 min at 1500g, washed with 70% ethanol and allowed to air dry.

Annealing Annealing of template and primer was carried out by resuspending the DNA in a 10ul volume containing 200mM Tris-Cl pH 7.5, 100mM $MgCl_2$, 250mM NaCl, 10% DMSO and 10ng of the relevant primer, and incubating at 37^{0}C for 20min.

Labelling Reaction 1ul 0.1M DTT, 1ul 1.5uM each dNTP (Boehringer), 2 units Sequenase enzyme (USB) and 5uCi of ^{35}S-labelled nucleotide triphosphate (>1000 Ci/mM; Amersham) were added to the annealing reaction and incubated for 5min at room temperature. The labelled nucleotide triphosphate was chosen according to the composition immediately 3' to primer binding site.

The termination step was carried out by aliquoting 3.7ul of the labelling reaction into each of 4 termination mixtures, each containing 80uM of the 4 dNTPs, 8uM of either (dideoxy-)ddGTP, ddATP, ddTTP or ddCTP, contained within a 2.5ul volume. Reactions were allowed to proceed at 37^{0}C for 5 min, before adding stopping the reaction by the addition of 4ul of loading buffer (95% formamide, 20mM EDTA, 0.05% bromophenol blue, and 0.05% cyanolFF).

Polyacrylamide-Gel Electrophoresis The sequencing reactions were resolved on 8% denaturing polyacrylamide gels, (BRL model S2 electrophoresis apparatus). Gels were prepared with 8% (w/v) 20:1 acrylamide:bisacrylamide (BDH), 0.1% (w/v) ammonium persulphate (Sigma) and 1xTBE. Gentle heat was applied to dissolve the reagents, and the solution allowed to cool. Prior to pouring the solution, TEMED (N,N,N',N'-tetramethylethylenediamine, Sigma) was added to 10%. Samples were heat denatured at 90^{0}C for 2min prior to loading onto the sequencing gel. Electrophoresis was carried out with 1xTBE as the running buffer, at 75 watts, until the bromophenol blue present in the loading buffer had egressed from the gel.

Following electrophoresis, gels were fixed with 5% acetic acid (BDH) and 5% methanol (BDH). The gels were dried under vacuum at 80^{0}C for 2 to 3hr, and autoradiography

carried out for 16 to 24hr, using Kodak XAR-2 Xray film.

Sequence Analysis Sequence analysis was performed on a VAX computer, using the University of Wisconson GCG package (Devereux *et al*, 1984).

2.9.2 Automated DNA Sequencing

Automated DNA sequencing was carried out using the model 373A DNA sequencing system of Applied Biosystems Inc.

The automated sequencing method, as with the manual sequencing, is based upon the di-deoxy chain termination method of Sanger *et al*, 1977. With the automated sequencing employed in this laboratory we use T7 DNA polymerase with dye-labelled terminators. The di-deoxynucleotides are labelled with a dye molecule each of which emits light at a different wavelength when excited by a laser which continually scans the gel while it is moving. A photon multiplier tube detects the fluorescent light and converts it to an electrical signal. The signal are transmitted to the computer and stored for eventual processing.

DNA sequencing of cloned material with an automated sequencer involves the following four steps:

- Amplification of the DNA by PCR, with a primer pair, of which, one is biotinylated.
- Purification of single-stranded DNA using a Dynabead biomagnetic separating system (Dynal AS).
- DNA sequencing.
- Data collection and analysis.

For the automated sequencing described in this study, the first three steps were carried out by Alexander Cleland (Centre for HIV Research, University of Edinburgh).

PCR Amplification

Single-stranded DNA is prepared for sequencing by the PCR-amplification of the required region with a primer pair of which one is biotinylated. The DNA strand primed from the biotinylated primer will incorporate the biotin at its 3' end.

Primers were specified to amplify a 1.7kb portion of the *env* gene, as three overlapping regions of 780, 690 and 510bp respectively. These are denoted 1, 2 and 3 in figure 2.2, respectively. The primers for the amplification and subsequent sequencing are as follows:

Region 1

Amplification:

3663 5' AGC AGA AGA CAG TGG CAA TG (+, 5773)

3664 5' TTC TGC TAG ACT GCC ATT (-, 6554)

Sequencing:

365F 5' GAT GTT GAT GAT CTG TAG (+, 5856)

0185 5' CAT GGC TTT AGG CTT TGA TC (-, 6108)

402 5' GAT CAA AGC CTA AAG CCA TG (+, 6126)

3664 as above

Region 2

Amplifcation:

332 5' TAC AAT GTA CAC ATG GAA TT (+, 6503)

405 5' CTT CTC CAA TTG TCC CTC ATA (-, 7192)

Sequencing:

332 as above (non-biotinylated)

405 as above

307 5' CTG GGT CCC CTC CTG AGG (-, 6861)

406 5' TCA GGA GGG GAC CCA GAA ATT (+, 6883)

Region 3

Amplification:

3893 5' GCT ATT AAC AAG AGA TGG (+, 7143)

3662 5' TTC TCT GTC CCA CTC CAT CC (-, 7653)

Sequencing:

3893 as above

3661 5' CAG CAG GAA GCA CTA TGG GC (+, 7363)

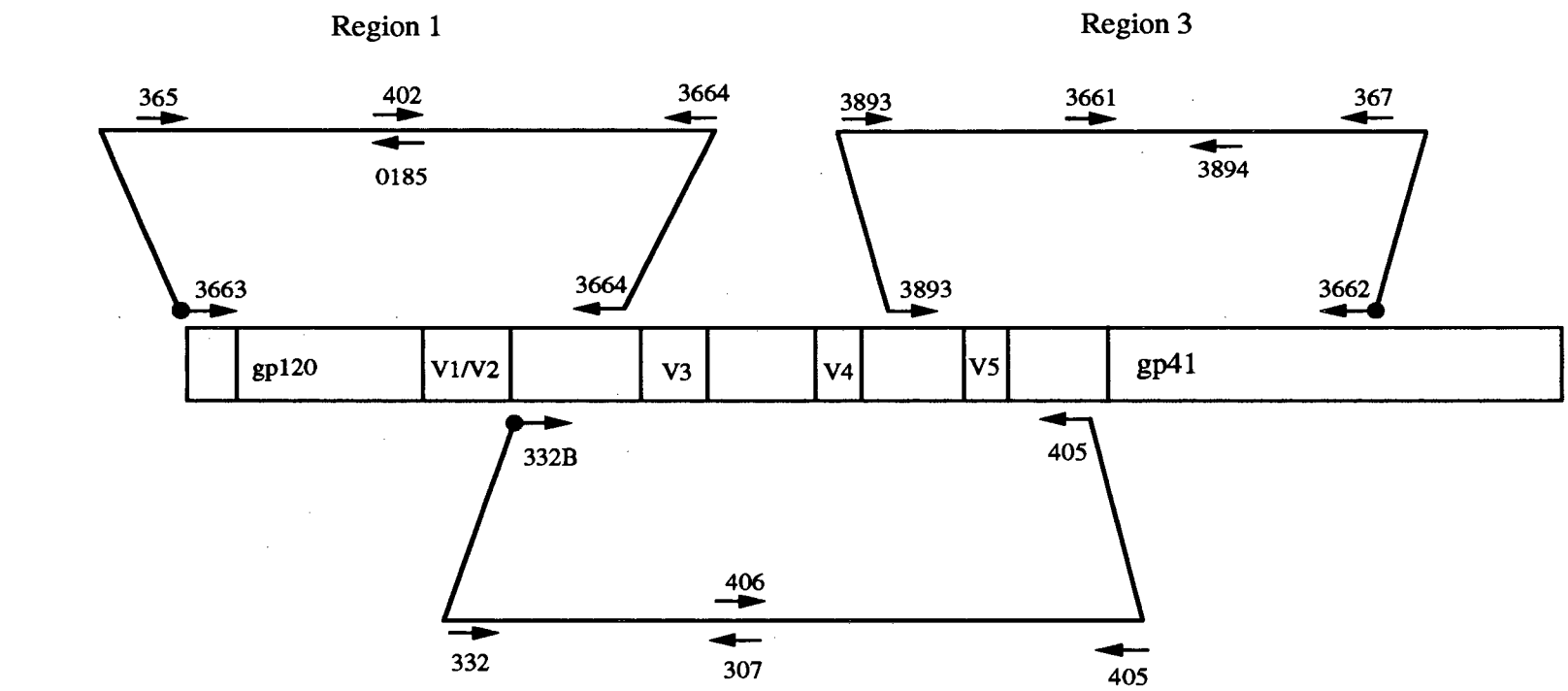
3894 5' CTG CTT GAT GCC CCA GAC (-, 7478)

367 5' GGA GTA ATA AAT GTC TAG ATG (-, 7630)

50ng of the appropriate DNA plasmid were amplified in a 50ul volume reaction, as described in section 2.7, with primers in molar ratio of 10pmol of biotinylated primer to 30pmol of non-biotinylated primer. Following the PCR, 10ul were analysed by agarose-gel electrophoresis to ensure the correct yield of DNA. The remaining 40ul of PCR product were taken forward into the two strands separated and purified as described below.

Preparation of Single-Stranded DNA For each sequencing reaction 20ul of dynabeads coated with streptavidin (Dynal AS) were prepared in 40ul of 0.1% bovine serum albumen and resuspended in 40ul of extraction buffer (10mM Tris-Cl pH7.5, 1mM EDTA, 20mM NaCl) to a final concentration of 5ug/ml. To this was added 40ul of PCR product, the solution mixed thoroughly and incubated at 48^0C for 30min. Following incubation, the reactions were placed in a magnetic separator (Dynal AS) and the supernatant discarded. The beads were then washed in 40ul extraction buffer and resuspended in 8ul 0.1M NaOH and incubated at room temperature for 10min to denature the non-biotinylated complementary strand. The reactions were then returned to the magnetic separator and the supernatant containing the unbound complementary strand was removed and neutralised by the addition of 4ul 0.2M HCl and 1ul 1M Tris-Cl pH7.5. The solution was then made up to 14ul with dH_20 and stored for subsequent sequencing. The beads containing the captured biotinylated DNA strand were washed three times, firstly, with 50ul NaOH, then 50ul extraction buffer, followed by 50ul TE buffer. The beads were then resuspended in 14ul dH_20.

DNA Sequencing The materials for the sequencing reactions were obtained from the 'PRISM Sequenase terminator' single-stranded sequencing kit (Applied Biosystems Inc). 14ul of single-stranded DNA template, prepared as described above, were mixed with 1ul of the appropriate sequencing primer (figure 2.2) at 0.8uM. To this was added 5ul of a freshly prepared 5x MOPS buffer (0.2M MOPS pH7.5, 0.25M NaCl and 0.05M $MgCl_2$, 25% glycerol, 25mM $MnCl_2$, and 0.075M isocitrate) were added. The reaction was incubated at 65^0C for 2min and allowed to cool slowly to room temperature.


Region 1
Region 3
365
402
3664
0185
3893
3661
367
3894
3663
3664
3893
3662
gp120
V1/V2
V3
V4
V5
gp41
332B
405
406
332
307
405
Region 2
biotinylated primer
non-biotinylated primer

4ul 'T7 Dye Terminator Mix' (as supplied in the sequencing kit) was added to the annealedtemplate reactions and pre-warmed at 37°C for 2min. After this time 1ul (1.5 units) of T7 DNA polymerase was added, the reaction mixed and incubated at 37°C for 10min. The reactions were then placed on ice and spun briefly to collect any condensation.

The excess dye was removed from the reactions containing the biotinylated strand by placing the tube in the magnetic separator, removing the supernatant and washing the DNA-bead complex twice with 50ul dye-terminator wash buffer (0.01M Tris-Cl pH8.0 and 0.1% Tween-20) and once with 50ul TE. The excess dye terminator was removed from the non-biotinylated strand by precipitating the DNA by the addition of 29.5ul 8M NH_4Ac, 150ul 95% ethanol. The precipitate was collected by centrifugation at 1500g for 20min, the supernatant removed and the pellet dried by heating to 90°C for 2min. Both the biotinylated and non-biotinylated samples were resuspended in 3ul formamide/EDTA solution (5:1 deionised formamide:50mM EDTA pH8.0). The sequencing products were resolved by electrophoresis on 6% polyacrylamide gels, within the 373A automatic sequencer, according to the manufacturer's instructions.

Collection and Processing of Sequence Data

The data from the sequencing run was collected on a Macintosh computer using the ABI data collection programme (ABI Inc).

Data in the form of ABI sequence files were transferred to a SUN Sparcstation computer via CAP (Columbia Appletalk Package). The data was processed through a preliminary editor TED (which forms part of the Staden computer package, see below), using the Seqprocess script (written by Chris Wade, Centre for HIV Research, Division of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3JN). The data were then entered into the Xbap data base (Staden package) using the Seqedit script (Chris Wade, address as above).

TED and Xbap are two programs that form part of the Staden Package, written by Roger Staden and Simon Dear, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH).

Sequences from both strands of the DNA, covering the entire region of gp120 sequenced from each clone, were aligned and contiguous sequences (contigs) assembled within the Xpab program. The full-length gp120 sequence was then edited by eye, with reference to the sequence trace data. The consensus gp120 sequence was output from Xpab and transferred to a VAX computer, where it was analysed using the GCG package. The sequencing data are collected and analysed by Macintosh IIci computer. Two programs are utilised Data Collection and Analysis. Sample files created from the preliminary analysis of data were transferred to a Sun Sparkstation computer and subsequently analysed using TED and Xbap.

2.10 *In Vitro* Transcription and Translation

In vitro transcription was carried out using 'phage T7 RNA polymerase, initiating transcripts from the T7 promoter sequence present in pBluescript (- located upstream of the *env* gene - see section 2.3 and chapter 3, section 3.6). *In vitro* translation was carried out using the following systems purchased from Promega: The Wheat Germ Extract System and the Rabbit Reticulocyte Lysates. The protocols followed were those recommended and detailed in the 'Promega Protocols and Applications Guide'. These will be summarised briefly below.

In Vitro Transcription

For each reaction 5ug of the appropriate plasmid construct were digested to completion with 10 units of *Xho*I (section 2.5). Following digestion the DNA was extracted once with phenol and once with chloroform and purified by ethanol precipitation. The DNA was resuspended in dH_20 and 2ug was used to direct *in vitro* transcription with T7 RNA polymerase (Promega). Each transcription reaction contained 1x transcription buffer (40mM Tris-Cl pH7.5, 6mM $MgCl_2$, 2mM spermidine and 10mM NaCl), 10mM DTT, 100u rRNAsin (Promega), 0.5mM rNTPs, 15U T7 RNA polymerase, 2ug DNA, made up to 100ul with nuclease-free dH_20. The reactions were incubated for 120mins at 37^0C.

Following incubation, the reactions were treated with DNaseI (Promega) at 1 unit per ug template DNA, by incubating the reaction at 37^0C for 15min. The RNA was purified

by phenol/chloroform extraction and ethanol precipitation. The RNA pellet was washed with 70% ethanol and allowed to air dry. The RNA was resuspended in nuclease-free dH_20 and stored at -70^{0}C.

In Vitro Translation

Prior to assembling the reaction components (either from the Wheat Germ Extract System, or the Rabbit Reticulocyte Lysates; Promega) the RNA was denatured by heating at 67^{0}C for 10min. Each reaction contained 25ul of wheat germ extract, 40units RNasin, 1mM amino acid mixture (minus methionine), 1M potassium acetate, 1ug RNA substrate, and 2.5ul [^{35}S]methionine (1,2000Ci/mmol; Amersham) at 10mCi/ml, and was made up to 50ul with dH_20. The reactions were incubated at 25^{0}C for 60min, and on completion were stored at -70^{0}C.

SDS Gel Analysis of Translation Products

Discontinuous gel electrophoresis under denaturing conditions in the presence of 0.1% SDS, a method described by Laemmli (1970), was carried out. Vertical slab minigels with a 12% SDS polyacrylamide separating gel and a 5% stacking gel were prepared for the Mighty Small II SE250 (Hoefer Scientific Apparatus).

5ul of wheat germ translation product were prepared for electrophoresis by adding 20ul loading buffer (containing 10% glycerol, 10% SDS, 2.5% B-mercaptoethanol), boiling for 5min and centrifugation for 5min. 10ul were loaded onto the gel along side ^{14}C-labelled Rainbow protein markers (Amersham) ranging from 14,000 to 200,000 daltons. Electrophoresis was carried out at 10mA whilst the samples were migrating through the stacking gel, and at 20mA when samples were migrating through the separating gel.

Once the bromophenol blue present in the loading buffer had reached the end of the gel, electrophoresis was stopped and the gel removed and fixed in 30% methanol, 10% glacial acetic acid, and stained with Coomassie Blue for 30min. After a series of destaining in 30% methanol and 10% acetic acid, the gel was dried onto a sheet of Whatman 3MM paper, using gel drying apparatus. Autoradiography was carried out using 'AGFA curix RP1' X-ray film (Kodak), exposing the film for 12-16hr.

2.11 Cell Culturing

The following two sections, 2.11 and 2.12, describe the experimental procedures for cell-culturing and the recovery of recombinant viruses. These were developed and carried out by Dr Robert Walker, with assistance from Mrs Elizabeth Harvey.

Cell Lines

COS-1 - an established simian cell line transformed with the simian virus SV40 (Gluzman, 1981). This cell line has a single integrated copy of the complete early regions of SV40 DNA and expresses the SV40 transforming T-antigen. This cell line was a gift from Dr Myra McClure, London.

The following cell lines were obtained from the MRC AIDS Reagent Project Repository, National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Herts EN6 3QG, UK:

C8166 - a CD4-positive T cell line

This cell line derives from the transformation of human umbilical cord blood lymphocytes with Human T-Cell Leukemia Virus type 1 (HTLV-1) (Salahuddin *et al*, 1983).

H9 - a CD4-positive T cell line

The H9 cell line is a clonal derivative of a T cell line designated HT (Mann *et al*, 1989). HT is a subline of the continuous T cell line HUT-78, established from a human cutaneous T-cell lymphoma (Gazdar *et al*, 1980).

Sup-T1 - a CD4-positive T cell line

This cell line was established from a non-Hodgkin's T-cell lymphoma (Smith *et al*, 1984).

Jurkat$_{tatIII}$ - a CD4-positive T cell line, expressing the HIV-1 *tat* gene

The Jurkat cell line, established from a Human T-cell leukaemia, has been transformed with the HIV-1 *tat* gene. This cell line is reported to support the growth of poorly

replicating primary HIV-1 strains (Fenyo *et al*, 1988).

U937 - an histiocytic cell line with monocyte-like characteristics

This cell line derives from an histiocytic lymphoma, and phenotypic characteristics of the histiocyte-monocyte series (Sundstrom and Nilsson, 1976).

Isolation and Culture of Peripheral Blood Mononuclear Cells

Whole blood was obtained from normal healthy HIV-1 seronegative donors. Peripheral blood mononuclear cells (PBMCs) were isolated by density centrifugation (400g for 30min) on sodium metrizoate/Ficoll (Lymphoprep; Nycomed Limited, Birmingham, UK) and washed twice in RPMI medium (HyClone Europe Ltd, Cramlington, UK). PBMCs were resuspended at $2x10^6$cells/ml in RPMI-complete medium; (RPMI supplemented with 2mM L-glutamine (HyClone), penicillin (50U/ml; GIBCO BRL, Paisley, Scotland), streptomycin (50ug/ml; GIBCO) and 10% foetal calf serum (FCS; Advanced Protein Products Ltd, Brierley Hill, UK). Cultures were stimulated with phytohaemagglutinin (PHA; 5ug/ml; Sigma Chemical Company Ltd, Poole, UK) for 72hrs and then maintained in RPMI-complete medium further supplemented with 30U/ml recombinant human interleukin-2 (IL-2; Advanced Protein Products).

Maintenance of Cell Lines

The immortalized $CD4^+$ T cell lines; H9, C8166, SupT1, Jurkat tat_{III}, and MT2 and the myeloid cell lines: U937, HL60 and THP1 were maintained in RPMI-complete medium and sub-cultured twice weekly. COS-1 cells were maintained in Dulbecco's Modified Eagle's medium (DMEM) supplemented as with RPMI-complete medium.

2.12 Transfection and Culture of Recombinant Virus

Preparation of DNA

DNA of full-length proviral clones were prepared for transfection from small-scale (3ml) cultures of *E.coli*, using commercially available DNA purification columns, either

from Promega (the 'Magic Miniprep' purification system), or plasmid purification columns from Qiagen. These systems allow the preparation of small amounts of DNA (3-5ug) of sufficient purity for transfection, without resorting to conventional purification by caesium chloride-gradient centrifugation, which would have resulted in the handling of larger amounts of potentially infectious material.

For each transfection, 2ug of plasmid DNA from the purification columns was sterilised by ethanol precipitation. The DNA was collected and washed with 70% ethanol and allowed to air-dry under sterile conditions (by storage within a Class II biosafety cabinet). The DNA sample was resuspended in 21ul sterile dH_2O. 1ul was assayed by agarose-gel electrophoresis to ensure there was efficient recovery of the DNA following the ethanol precipitation.

To recover virus from the two molecular clones $SF2_{MC}$ and $SF162_{MC}$, 5ug of each DNA from the 5' and 3' subclones was linearised by digestion with *Eco*RI and co-precipitated with ethanol. The DNA pellet was air-dried and resuspended in sterile distilled H_2O.

DNA Transfection

Cos cells ($1\text{-}2x10^5$) were seeded into 6 well plates (Nunc, GIBCO) 24hrs prior to transfection. Plasmid DNA (5-10ug) in 100ul OPTI-MEM medium (GIBCO) was mixed with lipofectin reagent (GIBCO; 15ug in 100ul OPTI-MEM) for 15min at room temperature (RT). COS-1 cell monolayers were gently washed twice with OPTI-MEM and overlaid with 0.8ml OPTI-MEM. Lipofectin-DNA mixture was added to the COS-1 cells and incubated for 6hrs at 37°C, 5% CO_2, after which 2ml DMEM-complete medium (20% FCS) was added. Alternatively, cell suspensions ($3x10^6$ cells in 0.8ml OPTI-MEM) of immortalized $CD4^+$ T cells or myeloid cells were combined with the lipofectin-DNA mixture as above and subsequently cultured in RPMI-complete medium (20% FCS). After 24hrs, $5x10^6$ PHA stimulated PBMCs or $3x10^6$ cells from transformed T cell and myeloid cell cultures were co-cultivated with transfected COS-1 cell monolayers to enable spread and amplification of any resultant recombinant virus progeny. 72hrs later PBMCs and transformed cells of co-cultures were transferred to T25 flasks and monitored for viral production. Additional PHA stimulated PBMCs or appropriate cells from immortalized cell cultures were added to transfected cultures as

required.

Assessment of Viral Production

Transfected cultures were monitored twice weekly for 4-6 weeks post-transfection for production of virus. At each time, 1ml of culture was clarified by centrifugation (250g for 10min) to obtain the supernatant which was filtered (0.45um) and stored at -70°C for subsequent determination of p24 antigen levels by an enzyme-linked immunosorbent assay (ELISA). Recovered cells were stained for *gag* core antigens by direct immunofluorescence.

Direct Immunofluorescence on Fixed Cells

Cell pellets were washed twice in phosphate buffered saline (PBS; pH7.5) and permeabilized by hypotonic shock (PBS:distilled H_2O; 1:1). Duplicate aliquots (15ul) were spotted onto PTFE coated multiwell microscope slides (C.A.Hendley Ltd, Loughton, UK) and allowed to air dry for 45mins. Slides were fixed in acetone:methanol (1:1) for 10min at RT. Cell preparations were stained with FITC-conjugated mouse anti-HIV-1 p24 monoclonal antibody (MAb; Coulter Clone KC57-FITC, Coulter Immunology, Luton, UK) and an isotype control (Coulter Clone MsIgG1-FITC). MoAbs were diluted 1:200 in PBS containing 1% bovine serum albumin (Sigma) and incubated for 45min at RT. Slides were washed in PBS (3 changes over 10min period) and briefly (30sec) counterstained in 0.05% Evans Blue (BDH Laboratory Supplies, Glasgow, Scotland). Following mounting in glycerol containing an anti-fading agent (Citifluor, Canterbury, UK) slides were examined under a Nikon OPTIPHOT microscope with episcopic-fluorescence attachment (Nikon Ltd, Kingston, UK).

p24 Antigen ELISA Supernatants were assayed for p24 antigen using affinity purified sheep anti-p24 antibody as capture antibody (Aalto BioReagents, Dublin, Eire), biotinylated mouse anti-p24 MAb as detector antibody (MRC AIDS Directed Programme) followed by streptavidin-horseradish peroxidase (Sigma) and o-phenylenediamine dihydrochloride (OPD; Sigma) for chromogenesis. Briefly, 100ul capture antibody (1mg/ml) was diluted 1:100 with coating buffer (150mM $NaHCO_3$,

pH9.0) and 100ul dispensed to 96 well high binding microelisa plates (Costar, High Wycombe, UK) and incubated overnight at RT. After washing in Tris buffered saline (TBS; 144mM NaCl, 25mM Tris, pH7.5) plates were blocked for 30min at RT using 200ul/well of 2% Marvel milk solution, then washed in TBS. Following heat inactivation (56°C for 45min) viral lysates prepared from culture supernatants treated with Empigen (0.1% final concentration; Calbiochem, La Jolla, CA), and p24 standards (1-1000ng/ml; Baculovirus-derived p24, American Bio-Technologies Inc, Cambridge, MA) were added to plates and incubated overnight at RT. Plates were washed in TBS/Tween (0.05%; Sigma) and 100ul biotinylated mouse anti-p24 MAb (diluted 1:1000 in TBS/0.05% Tween/20% FCS) added and incubated at RT for 2hr. Subsequent addition (100ul) and incubation (1hr at RT) of Extravidin-horseradish peroxidase(diluted 1:4000 in PBS/0.05% Tween) followed washing in TBS/0.05% Tween. A final wash in TBS/0.05% Tween preceded addition (100ul) and incubation (15-20min) of OPD (1mg/ml in 0.05M phosphate citrate buffer with urea, pH5.0; Sigma). Reaction was stopped by addition of 100ul 5N H_2SO_4. Absorbances at 492nm were read using a Multiskan Bichromatic microplate reader (Life Sciences International Ltd, Basingstoke, UK) and data analysed using GENESIS software package (Life Sciences).

Viral Stock Production and Infectivity Determination Cells from p24 antigen positive cultures were mixed 1:4 with uninfected cells from $CD4^+$ T cell lines for 72hr. Supernatants were harvested following centrifugation (300g for 10min), filtered (0.45um), mixed 1:1 with FCS and stored in liquid N_2. The infectivities of supernatants were determined by performing sextuplet replicates of 4-fold serial dilutions of virus (50ul) and incubating with 50ul of C8166 cells ($4x10^5$/ml) in 96 well flat bottomed microtitre plates (Costar) at 37°C, 5%CO_2 for 7-10 days. Plates were scored for the formation of syncytia and the 50% tissue culture infective dose ($TCID_{50}$) calculated by the Reed & Muench method.

Chapter 3

Construction of an HIV-1 Cassette-Vector (pHXB2-MCS)

3.1. Introduction

It is possible to produce infectious HIV-1 virions by the introduction of full-length, proviral clones into cell cultures (Levy *et al*, 1986, Adachi *et al*, 1986). This property may be exploited to examine the function of viral genes, or specific regions of the viral genome. Through the use of recombinant DNA techniques, specific genes, or genomic regions of the proviral genome, may be mutated, or exchanged for different allelic versions. By examining the phenotype of the resultant viruses, the function of the mutated gene, or exchanged genomic region, may be determined (Fisher *et al*, 1985, and 1986).

There have been many reports describing the construction of HIV-1 recombinant molecular clones for purposes of mapping the determinants of viral phenotype. (For example: Fisher *et al*, 1985; Cheng-Mayer *et al*, 1990; Groenink *et al*, 1991; Hwang *et al*, 1991). These were carried out by exchanging regions of the *env* gene between infectious molecular clones with differing phenotype, or by exchange of regions of the *env* gene from non-infectious clones into the backbone of an infectious clone. In these reports, the regions exchanged were limited to restriction-fragments common to both clones.

However, the region of interest may not be bounded by restriction sites shared between the clones. In such cases, it is possible to engineer additional restriction sites into the cloned sequence, for the exchange of more precisely-defined regions. The modified infectious clone may then be thought of as a 'cassette-vector', in which a defined region of the genome, that is, the 'cassette', can be replaced with variant sequences from another source. Indeed, it is possible to take advantage of the polymerase chain reaction (Mullis *et al*, 1987, Saiki *et al*, 1988), and design primers at the regions of the newly-specified restriction sites, that is, the cassette-boundaries, and so amplify sequences from a variety of sources, both cloned, and uncloned sequences, in a form in which they can be easily substituted into the vector. For example, there have been several reports of the engineering of infectious molecular clones to allow exchange of the V3-domain of gp120 - to study cellular tropism, cytopathic effect and neutralisation properties (De Jong *et al*, 1993a; Chesebro *et al*,

as the V2-loop, and an epitope in the external domain of gp41 (Fung *et al*, 1992, Muster *et al*, 1994).

The majority of these epitopes are conformationally-dependent - the neutralising sensitivity of the epitope is influenced by both the primary sequence and the conformation of the epitope, which in turn, can be influenced by amino acid changes distant to the epitope.

The conformational nature of these epitopes is seen clearly when neutralising-antibody escape mutants are found to map to regions of the protein that are outside the defined epitope. For example, antibody-escape mutants to the V3-loop epitope have been found to map outside the V3 loop - in some cases mapping to distant sites in gp120, and in gp41 (McKeating *et al*, 1989; Back *et al*, 1993). This is similarly the case for the complex CD4-binding site epitope, where resistant mutations are found to map outside this site (McKeating *et al*, 1993; Thali *et al*, 1994).

Ideally, to reproduce an HIV-1 envelope protein with the correct antigenic properties requires both gp120 and gp41 subunits of the envelope protein to be expressed, so they can assemble into their oligomeric, native form (Moore, 1993). This argues for the exchange-region of the proviral vector to contain the whole of the *env* gene.

However, another major factor influencing the design of a cassette-vector, where one sequence is to be replaced by a variant of it, is the organisation of the HIV-1 genome, which consists of many overlapping reading-frames (figure 3.1). The result of such overlapping reading frames is that the replacement of the coding sequence of one gene will result in the alteration of the coding capacity of any overlapping genes.

In this way, any resulting changes in given phenotype conferred by sequence changes in one gene cannot be attributable solely to that gene, the contribution of changes to the overlapping genes must be considered.

Figure 3.1. The Proviral Clone pHXB2-D

R *Eco*RI; S *Sal*I; B *Bam*HI; Xh *Xho*I; Xb *Xba*I

a - start *env* coding region; b - end of *vpu* coding region;

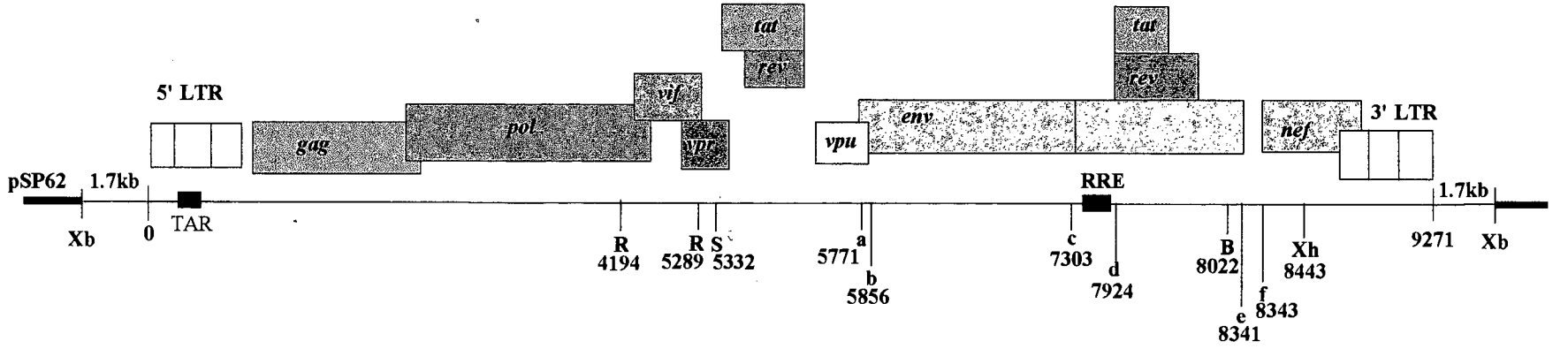
c - gp160 cleavage site; d - start of *tat* and *rev* 3' coding exon;

e - end of *env* coding region; f - start of *nef* coding region

co-ordinates according to HXB2R

opposite page 101

pHXB2-D



vector (pSP62) sequences

proviral sequences

cellular flanking sequences

The HIV-1 *env* gene overlaps with the *vpu* gene and the 3' coding exons of the regulatory genes *tat* and *rev*. To study the contribution to the sensitivity of the virus to neutralising antibodies, or indeed, to any other phenotypic properties, of changes limited to the *env* gene, the region exchanged should include only *env* gene sequences, and not those of *vpu*, *tat* and *rev*. Such a region lies in the intervening sequences between the overlapping reading frames of *vpu*, *tat* and *rev* (from nt 5876 to 7924 HXB2R). This region is 2.0kb and includes the sequence of the *env* gene encoding for the whole of gp120, and the majority of the extracellular portion of gp41. This region contains the major neutralising epitopes of gp120 and gp41.

A third region of the *env* gene was also considered as important for investigating the neutralising sensitivities of the virus envelope: the region of the *env* gene encoding for surface protein, gp120. The main reasons for exchanging gp120 alone is to prevent the formation of a hybrid gp41, and to exclude from the exchange region the Rev response element (RRE) which lies within the *env* gene (figure 3.1). The RRE is the *cis*-acting element for the viral regulatory protein Rev (Malim *et al*, 1990). It is conceivable that different RRE sequences contained within exchanged variants of the *env* gene may not function optimally with the HXB2 Rev protein. Omitting the RRE from the region exchanged should control for this type of variation.

In summary, we would like to design a cassette-vector with three exchange-regions: i) the whole of the *env* gene (from 5771 to 8341 HXB2R); ii) the region of the *env* gene lying between the overlapping coding sequences of *vpu*, *tat* and *rev* (from 5856 to 7924 HXB2R); and iii) the region of the *env* gene encoding for gp120 but not for gp41 (from 5771 to 7303 HXB2R).

The proviral clone, pHXB2-D, contains a number of unique restriction sites which could act as cloning sites (for example *Sal*I at 5332 , *Bam*HI at 8021, and *Nhe*I at 6806, and *Xho*I at 8443; figure 3.1, and appendix B). The *Xho*I site at 8443, provides a suitable 3' cloning site for the exchange of the entire *env* gene. However, there are no unique restriction sites at the 5' end of the *env* gene, nor are there sites at the 5' end of the overlapping reading frames of *tat* and *rev*, nor at the gp120-gp41 cleavage

site.

In consequence, three unique restriction-enzyme sites were engineered into the HXB2 sequence. A *Bst*EII site has been engineered at the 5' end of the *env* gene (at 5875), 114 nucleotides into the *env* gene. An *Xba*I site has been engineered at 7623, 300 nucleotides upstream of the *tat* and *rev* splice-acceptor site. An *Mlu*I has been engineered at 7279, 24 nucleotides upstream of the gp120-gp41 cleavage sites. In addition, a *Not*I site, at 5645, in the *vpu* gene, was also engineered into the HXB2 sequence, to act as an additional cloning site.

These unique restriction sites have been used to define three cassettes that can be readily exchanged for homologous sequences amplified from patient material:

- **cassette-1**: from the *Bst*EII site to the *Xba*I site (5875 to 7624; figure 3.2). This is a 1.7kb fragment of the *env* gene, encompassing the whole of gp120 (minus six highly conserved amino acids at the N-terminal), and the first 300 amino acids of gp41, including approximately three quarters of the external domain. This fragment lies entirely between the overlapping reading frames of *tat* and *rev*, and as such, the exchange of this fragment will not alter the coding capacity of these genes.

- **cassette-2**: from the *Bst*EII site to a naturally occurring *Xho*I site in the *nef* gene (5875 to 8443; figure 3.2). This is a 2.5kb fragment of the proviral genome encompassing the whole of the *env* gene. Although this fragment encodes for the whole envelope protein, exchange of this fragment will also result in changes to the *tat*, *rev* and *nef* genes, which may have a contributory effect to the viral properties *in vitro*.

The *Xho*I site, defining the 3' boundary of this cassette-region is in a highly conserved region of the *nef* gene, and in consequence, is found in 80% of isolates found in the Los Alamos data base. It is, in addition, rare to HIV-1 genomes, and therefore unlikely to be found in the *env* genes amplified from patient-material.

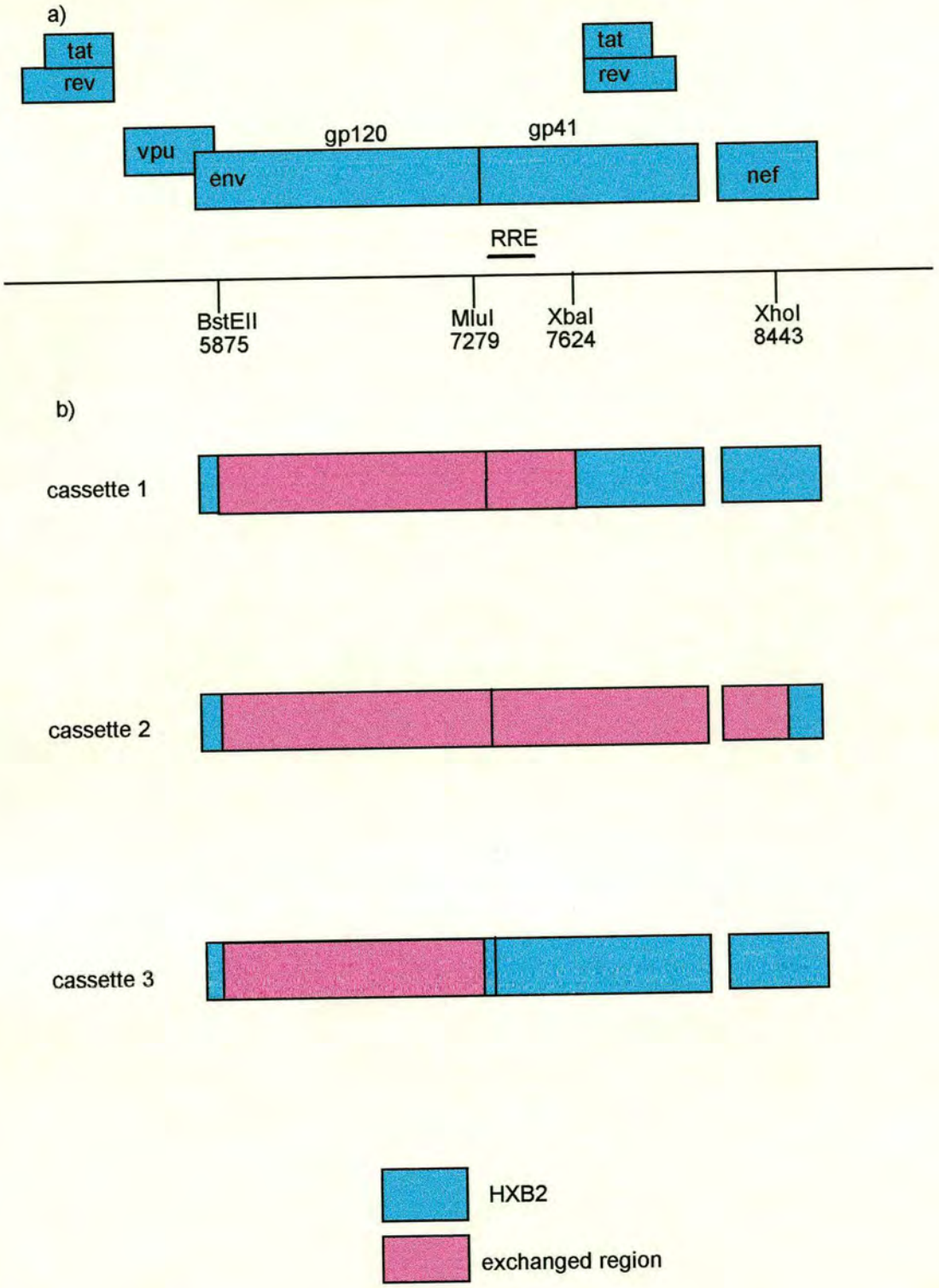
Figure 3.2. The Cassettes Specified in pHXB2-MCS.

a) The structure of the 3' half of the HXB2 genome, showing the open-reading frames and the location of the introduced-restriction sites.

b) The structure of the three cassettes in relation to the open-reading frames of the *env* and *nef* genes.

The boxes indicate open-reading frames; the blue boxes indicate the open-reading frames encoded by HXB2, and the pink boxes indicate those parts of the open-reading frames that will be exchanged on replacement of one of the three cassettes.

opposite page 104


a)
tat
rev
tat
rev
vpu
env
gp120
gp41
nef
RRE
BstEII
5875
MluI
7279
XbaI
7624
XhoI
8443
b)
cassette 1
cassette 2
cassette 3
HXB2
exchanged region

- **cassette-3**: from the *Bst*EII site to the *Mlu*I site (5875 to 7279; figure 3.2). This is a 1.4kb fragment of the *env* gene, encoding for all of gp120 (minus 6 highly conserved amino acids at the N-terminal and five, highly conserved amino acids upstream of the gp120-gp41 cleavage site). Exchange of this region will therefore produce a recombinant envelope protein, whose gp120 is encoded by sequences obtained from the patient, and whose gp41 encoded by the vector. In addition, the Rev-responsive-element which lies within the gp41 part of the *env* gene, in constructs generated by exchange of this region, will be encoded by the vector, and therefore acted on by its own Rev protein.

The *Bst*EII and *Xba*I site are each one synonymous nucleotide change away from the HXB2 sequence, and the *Mlu*I site is two synonymous nucleotide changes away from the HXB2 sequence (table 3.1). As a result of this, the engineering of these nucleotide changes into the HXB2 sequence, should not alter the coding capacity of the vector. In addition, each of these sites is rarely found in the *env* gene of diverse HIV-1 isolates, (as determined by examining the sequences found in the Los Alamos data base of HIV-1 sequences (Myers *et al*, 1991). As such, these sites are unlikely to be found in the *env* genes of sequences amplified from patient material; the presence of an internal site for one of these enzymes would render such sequences extremely difficult to clone into the vector.

The changes from the HXB2 sequence required to generate the *Not*I site result in two conservative amino acid changes of valine to alanine in *vpu* gene (table 3.1). However, these changes are unlikely to effect the growth properties of the resultant virus, as the *vpu* gene is not expressed in HXB2 due to the absence of a methionine start codon (Schwartz *et al*, 1990b).

In this study, the majority of recombinant clones were generated through the exchange of cassette-1, because of the practical constraints of amplifying sequences from PBMC-DNA. It was found to be relatively straight-forward to amplify by PCR the 1.7kb fragment - the region defined by cassette-1, but it was extremely difficult to amplify cassette-2 which is of 2.5 kb in length.

Table 3.1. Nucleotide Changes made to pHXB2-D in the Construction of the Cassette-Vector, pHXB2-MCS.

*Bst*EII at 5875 G'GTCACC A -> C at 5881	G G T C A C A G G T C A C C V T	HXB2 MCS amino acid
*Xba*I at 7624 T'CTAGA G -> A at 7627	T C T C T G G A A T C T C T A G A A S L E	HXB2 MCS amino acid
*Mlu*I at 7279 A'CGCGT A -> C at 7280 A -> C at 7282	A A G A G T G A C G C G T G R V	HXB2 MCS amino acid
*Not*I at 5645 GC'GGCCGC T -> C at 5645 A -> G at 5646 T -> C at 5648 A -> C at 5649	G T A G T A G C V V G C G G C C G C A A	HXB2 amino acid MCS amino acid

Co-ordinates according to HXB2R (Myers *et al*, 1991)

3.3. Cloning Steps Undertaken to Construct the Cassette-Vector pHXB2-MCS.

The manipulations carried to create the cassette-vector pHXB2-MCS from pHXB2-D are given in outline in figures 3.3(a) and (b). A more detailed representation of pHXB2-D and pHXB2-MCS may be found in figures 3.1 and 3.5, respectively.

Firstly, in order to carry out site-directed mutagenesis to introduce the three restriction sites, the *Eco*RI to *Xho*I fragment (5289 to 8443) of pHXB2-D was subcloned into the plasmid vector pBluescriptII(KS-) at the *Eco*RI and *Xho*I sites in the vector polylinker. The resultant clone, designated pRX, contains the entire *env* gene coding sequence and, in addition, the coding sequences of *tat*, *rev* and *vpu*. The three unique restriction sites for *Not*I, *Bst*EII and *Xba*I were introduced sequentially into pRX by oligonucleotide-directed mutagenesis (chapter 2, section 2.8).

Prior to replacement of the HXB2 *env* gene with the modified sequence from pRNBXX, it was first necessary to transfer the *Xba*I proviral insert fragment of pHXB2-D into a different cloning vector. The proviral sequence of pHXB2-D is contained in the plasmid vector pSP62 (Melton *et al*, 1984). However, this vector was unsuitable for constructing the cassette-vector, because it contains a number of 'unwanted' restriction sites; these include the sites chosen as cloning sites for the cassettes, *Xho*I, *Xba*I, and others which may be useful for other manipulations of the HXB2 sequence, such as *Sal*I which has a single site in HXB2. The sequence of pSP62 is not available on the database, which confounds the removal of these sites by conventional genetic manipulations.

Therefore, the HXB2 provirus and cellular flanking sequences were transferred to another vector, pSPTBM20Δ. The manipulations involved are shown schematically in figure 3.4. The *Xba*I fragment of pHXB2-D (which includes the entire provirus and cellular flanking sequences) was isolated and cloned into the *Nhe*I site of the plasmid pSPTBM20Δ, a deletion-derivative of pSPTBM20. pSPTBM20Δ was created by removing the majority of the polylinker (including the *Xba*I, *Xho*I and *Sal*I sites), by

double-digestion with *Eco*RV and *Sma*I, followed by self-ligation. By cloning into the *Nhe*I site of pSPTBM20D, both the *Nhe*I and *Xba*I sites are lost. The resultant recombinant molecular clone, pHXB2-SPT, has an insert fragment which is identical to that of pHXB2-D, excepting the removal of the *Xba*I site at the cloning junction.

In the final step to create pHXB2-MCS, the *Sal*I-*Bam*HI fragment of pHXB2-SPT (5332 to 8443) was exchanged for the *Sal*I-*Bam*HI fragment of pRNBXX (figure 3.3(b)). In pHXB2-MCS, the restriction sites defining the cassette-boundaries (*Bst*EII, *Xba*I and *Xho*I) are unique (figure 3.5), thus allowing the direct exchange of these regions with sequences amplified from PBMC DNA, when amplified with primers specified to incorporate these restriction sites in frame with the *env* and/or *nef* genes (chapter 4).

The *Mlu*I site specified as the 3' boundary of cassette 3, was introduced into pHXB2-MCS by Dr Peter Balfe, at Middlesex Medical School, London, using the site-directed mutagenesis method described by De Jong *et al*, 1992b. This construct was used only towards the end of this study, where it was employed for the construction of a single recombinant clone (see chapter 4, section 4.2.2). As a consequence, this vector will be referred to as pHXB2-MCS.*Mlu*I to distinguish it from pHXB2-MCS.

Figure 3.3 A Schematic Representation of the Steps Involved in Engineering the Proviral Cassette-Vector pHXB2-MCS

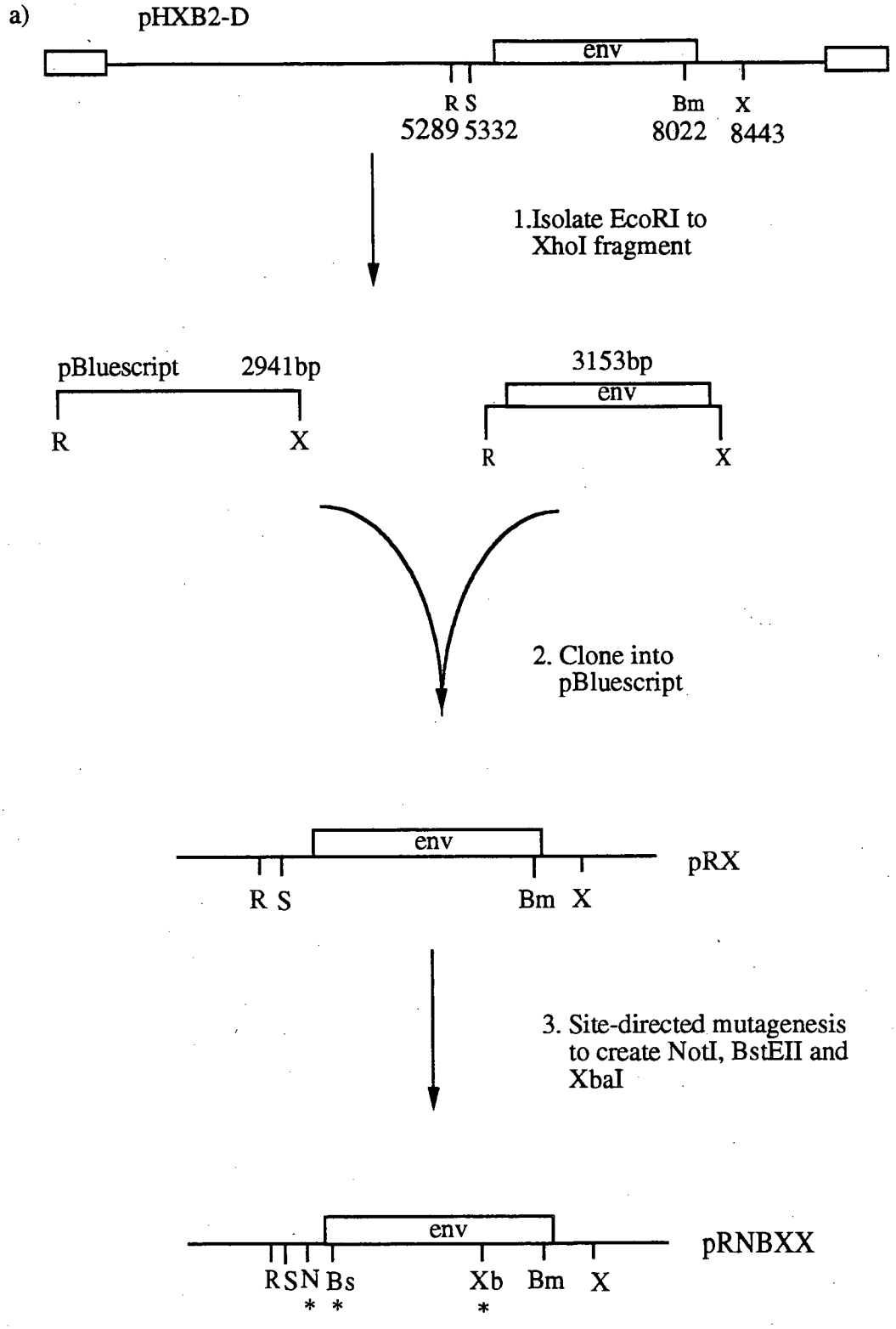
a) The subcloning of the *env* gene of pHXB2-D and the introduction of a *Not*I, *Bst*EII and *Xba*I site, and
b) the replacement of the *env* gene of pHXB2-SPT with the modified version.

The relevant DNA fragments are shown schematically and the restriction sites given. The *env* gene coding sequence and the proviral LTRs are indicated by open boxes.

The restriction sites are noted in abbreviated form:
R (*Eco*RI), S (*Sal*I), N (*Not*I), Bs (*Bst*EII), Xb (*Xba*I), Bm (*Bam*HI) and X (*Xho*I).

A more detailed representation of pHXB2-D and pHXB2-MCS may be found in figures 3.1 and 3.5, respectively.

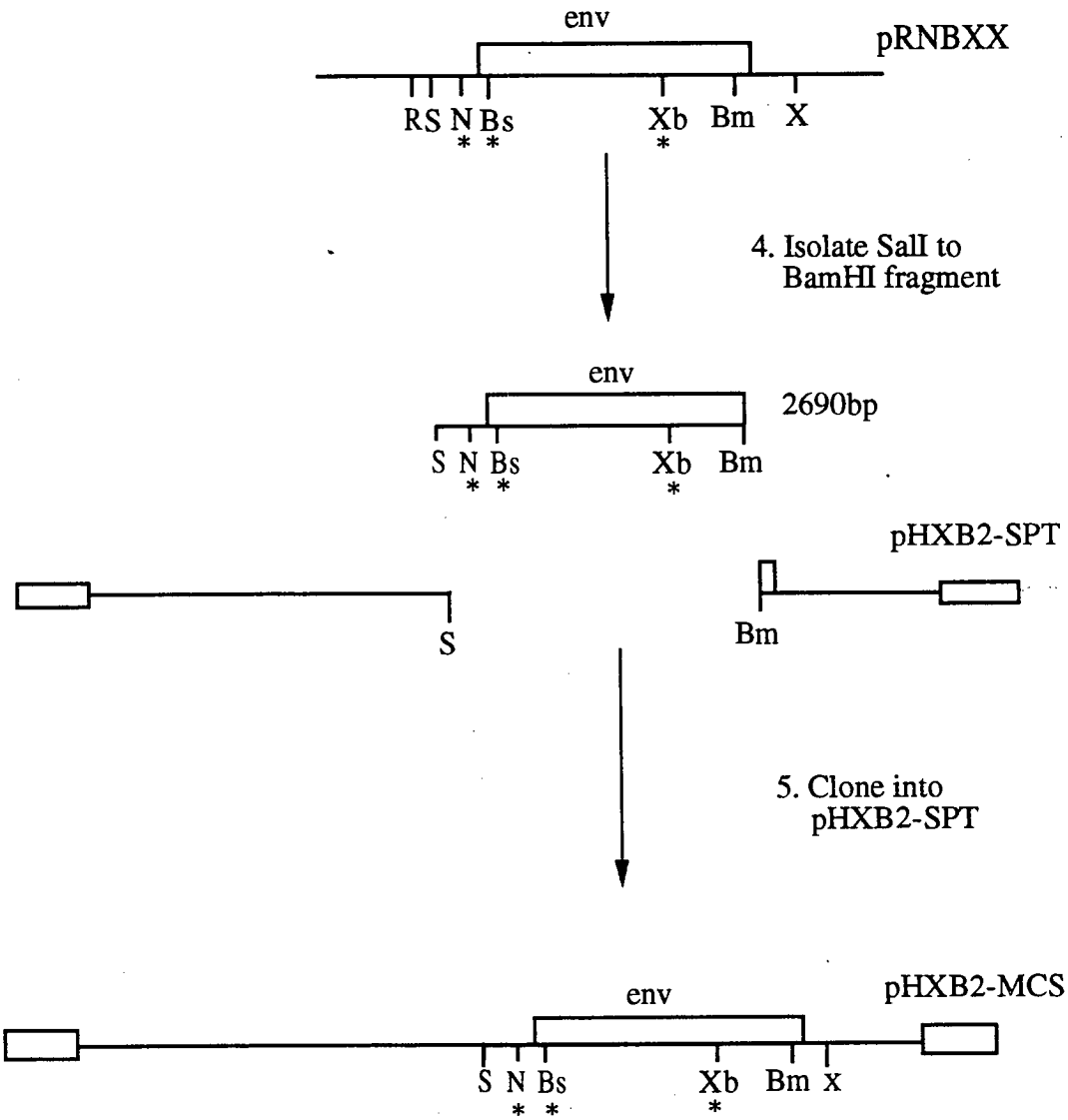
opposite page 109


a)
pHXB2-D
env
R S
5289 5332
Bm X
8022 8443
1.Isolate EcoRI to
XhoI fragment
pBluescript 2941bp
R X
3153bp
env
R X
2. Clone into
pBluescript
env
R S
Bm X
pRX
3. Site-directed mutagenesis
to create NotI, BstEII and
XbaI
env
R S N Bs
* *
Xb Bm X
*
pRNBXX
* synthetic
restriction site

continued.....

continued...

b)

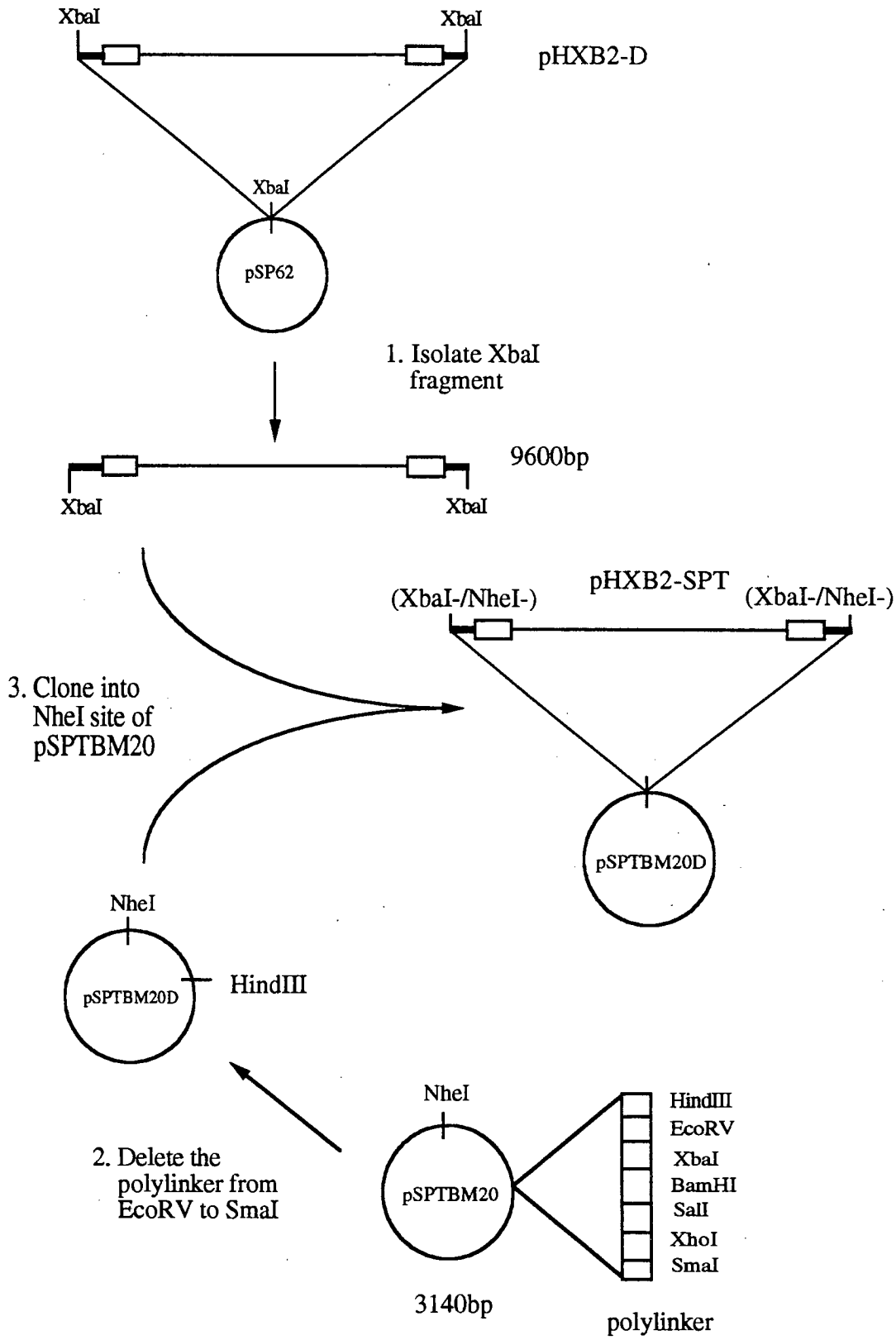


* synthetic
restriction site

Figure 3.4 A Schematic Representation of the Steps Involved in Generating pHXB2-SPT

The thin lines represent the proviral sequence, with thicker lines representing the cellular-flanking sequences. The proviral LTRs are shown as open boxes. Circles indicate the plasmid vector sequences.

opposite page 111


XbaI
XbaI
pHXB2-D
XbaI
pSP62
1. Isolate XbaI fragment
9600bp
XbaI
XbaI
pHXB2-SPT
(XbaI-/NheI-)
(XbaI-/NheI-)
3. Clone into NheI site of pSPTBM20
pSPTBM20D
NheI
pSPTBM20D
HindIII
NheI
pSPTBM20
HindIII
EcoRV
XbaI
BamHI
SalI
XhoI
SmaI
2. Delete the polylinker from EcoRV to SmaI
3140bp
polylinker

3.4. Construction of Vectors pHXB2-MCSΔ*env*, pNBXX and pNBXXΔ*env*

Three derivatives of pHXB2-MCS were generated in this study for cloning and expression purposes.

- **pNBXX** (5.7kb) - an *env* gene subclone of pHXB2-MCS. The *Not*I to *Xho*I fragment (5645 to 8443) of pRNBXX (figure 3.3) was subcloned into the plasmid vector pBluescript, at the *Not*I-*Xho*I sites in the vector polylinker (figure 3.6).

- **pNBXXΔ*env*** - (5.3kb) a deletion-derivative of pNBXX (figure 3.6), formed by removing the *Nde*I-*Stu*I fragment (5952 to 6380). This is a 428bp fragment of the *env* gene encoding the V1, V2 and C2 regions. It was removed by digesting pNBXX with *Stu*I and *Nde*I, gel-purifying the vector fragment, filling in the 3' overhangs with Klenow and self-ligating.

- **pHXB2-MCSΔ*env*** - a derivative of pHXB2-MCS, used as the cloning-vector to construct the full-length, recombinant, proviral clones of this study (figure 3.5). pHXB2-MCSΔenv was constructed by replacing the *Bst*EII-*Xba*I fragment of pHXB2-MCS, with that of the 428bp-deletion-containing *Bst*EII-*Xba*I fragment of pNBXXΔ*env* (see above).

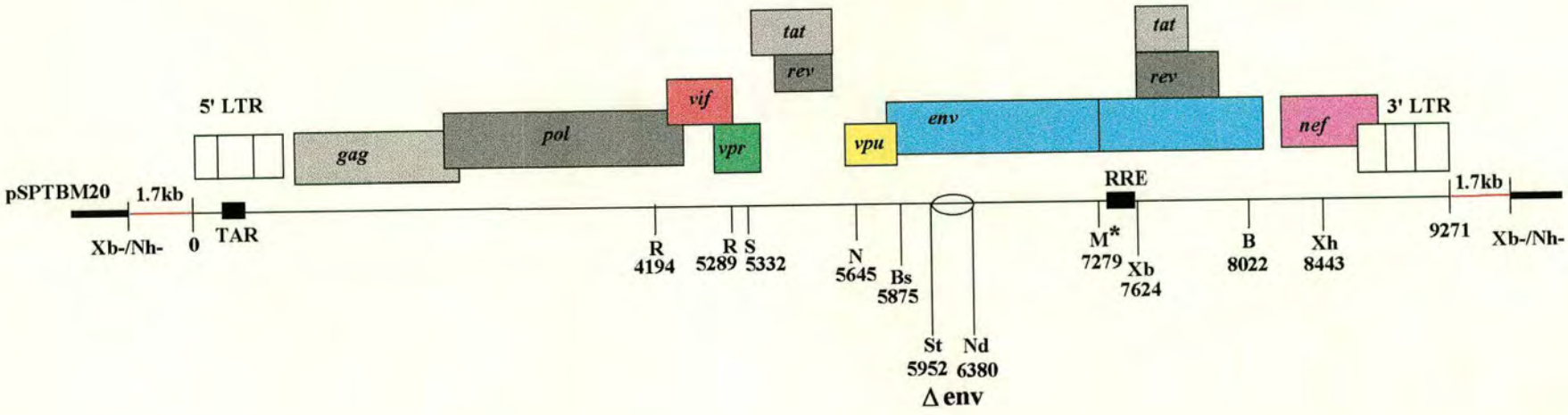
The 428bp deletion in the *env* gene of pHXB2-MCSΔ*env* allows the full-length, recombinant derivatives formed by substitution of one of the cassettes to be distinguished from it by size. In addition, pHXB2-MCSΔ*env* is a defective provirus, and therefore safer to handle.

Figure 3.5. The Proviral Clone pHXB2-MCS

R *Eco*RI; S *Sal*I; B *Bam*HI; Xh *Xho*I; N *Not*I; Bs *Bst*EII; St *Stu*I; Nd *Nde*I;
Xb-/Nh- *Xba*I & *Nde*I sites removed; M* *Mlu*I only present in pHXB2-MCS.*Mlu*I
Δ*env* fragment is deleted in pHXB2-MCSΔ*env*
co-ordinates according to HXB2R

opposite page 113

pHXB2-MCS



vector sequences

proviral sequences

cellular flanking sequences

Figure 3.6. The Cloning Vector pNBXX

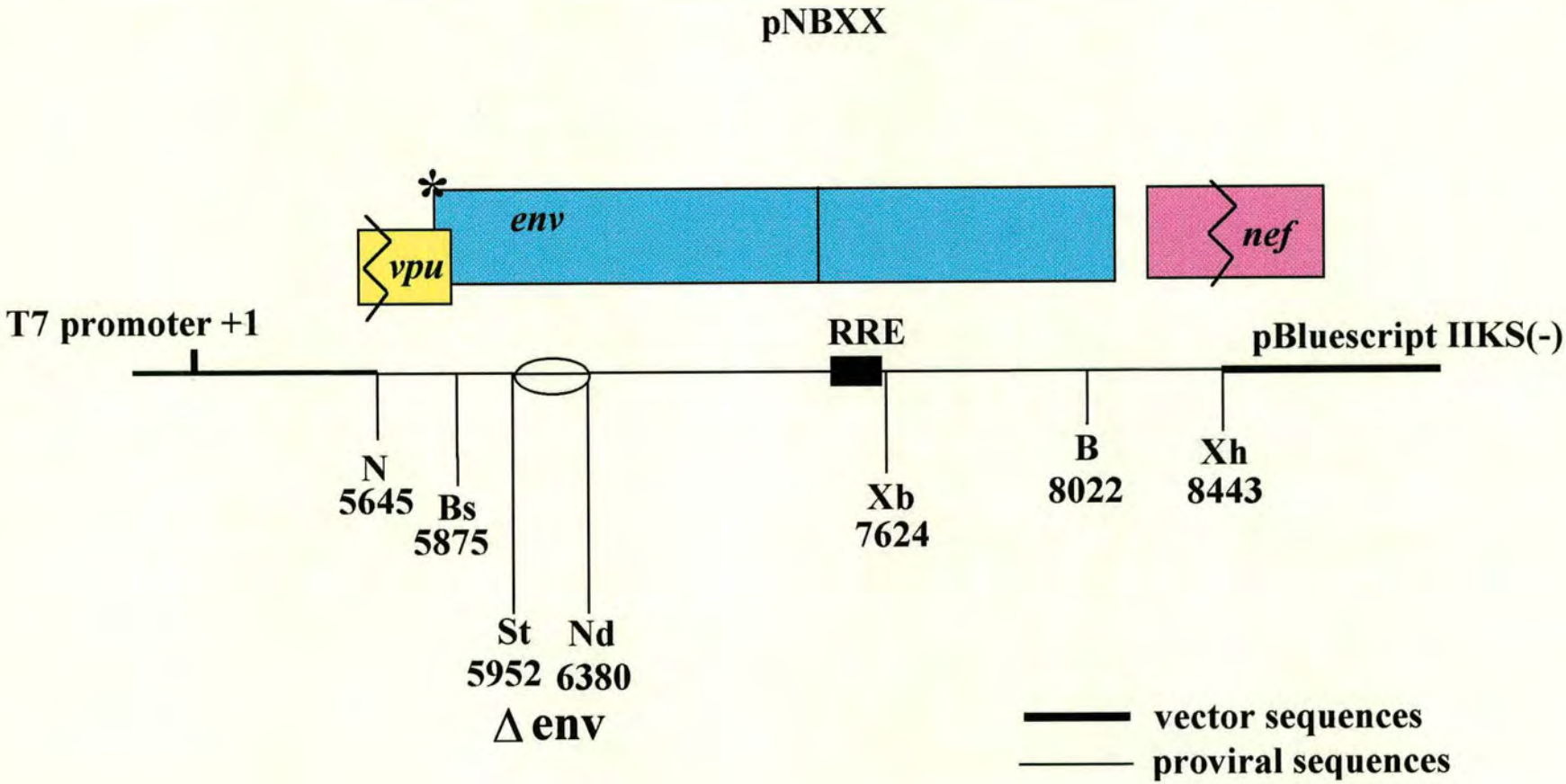
B *Bam*HI; Xh *Xho*I; N *Not*I; Bs *Bst*EII; St *Stu*I; Nd *Nde*I;

Δ*env* fragment is deleted in pHXB2-MCSΔ*env*

co-ordinates according to HXB2R

* env start codon

opposite page 114


pNBXX
*
env
vpu
nef
T7 promoter +1
RRE
pBluescript IIKS(-)
N
5645
Bs
5875
St
5952
Nd
6380
Δ env
Xb
7624
B
8022
Xh
8443
vector sequences
proviral sequences

3.5. Verification of the Sequence Changes made to Create pHXB2-MCS

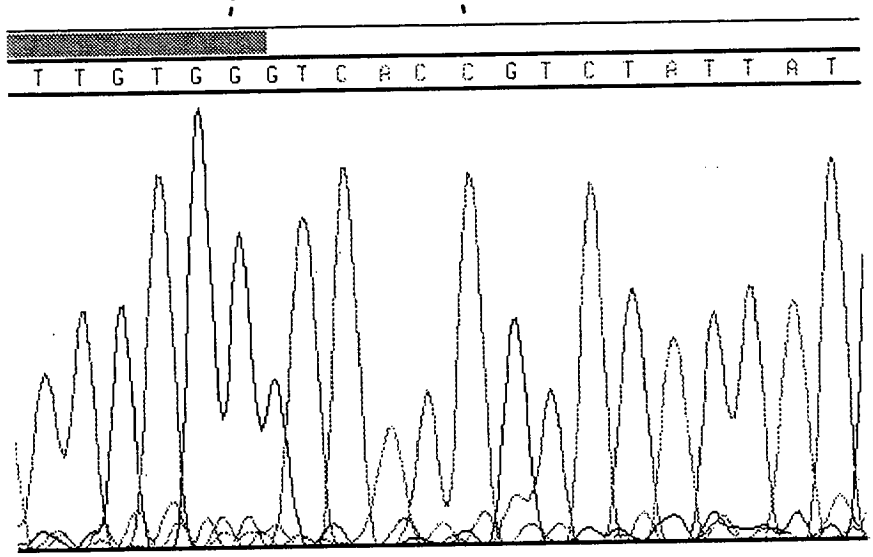
The incorporation of the three restriction sites, *Not*I, *Bst*EII and *Xba*I, into pHXB2-MCS, at the correct position in the proviral sequence was verified by extensive restriction mapping (data not shown). In addition, the nucleotide sequence changes made to create the *Bst*EII site and the *Xba*I site were verified by sequencing pHXB2-MCS at the site of the changes.

Solid phase sequencing and sequence-data analysis was carried out using an ABI automatic sequencing machine, as described in chapter 2, section 2.9.2. A copy of the chromatogram, showing the nucleotide sequence of pHXB2-MCS at the *Bst*EII and *Xba*I site, is given in figure 3.7a and 3.7b, respectively. At the *Bst*EII site there is a single A to C substitution from the parental (pHXB2-D) sequence, and at the *Xba*I site there is a single T to C substitution, with no other changes observed.

a)

GGTCACA - pHXB2-D

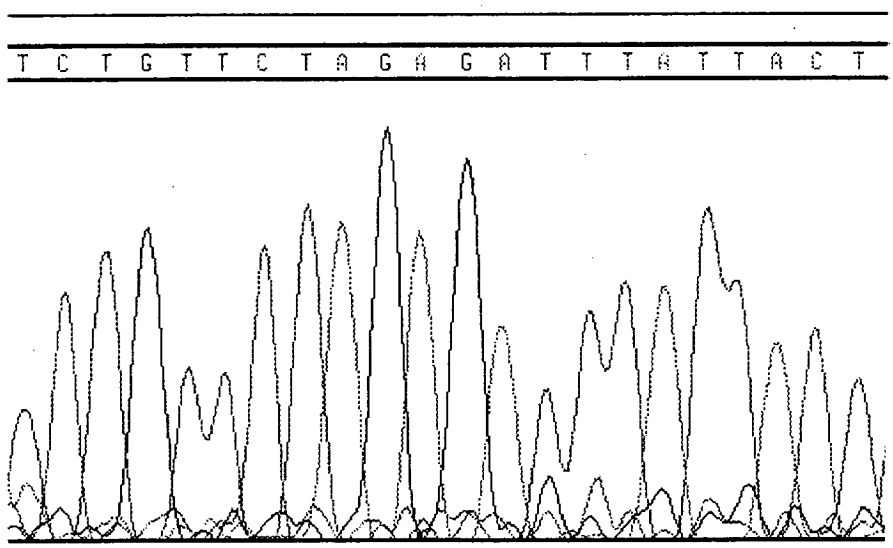
GGTCACC - pHXB2-MCS



b)

TCCAGA - pHXB2-D

TCTAGA - pHXB2-MCS



3.6. Verification of the Infectious Nature of pHXB2-MCS

Following the construction of pHXB2-MCS, the growth properties of virions produced from this clone were compared to the growth properties of virus produced from the unmodified clone, pHXB2-D. This work was carried out by Dr Jane McKeating, at the Chester Beatty Laboratories, London.

Three proviral clones were used in these experiments: pHXB2-SPT, pHXB2-MCS and pNL4-3, which is an additional clone derived from the HIV-1_{IIIB} isolate (Adachi *et al*, 1986). Virus was recovered by DNA-transfection into the CD4-positive transformed cell line H9, as described in McKeating *et al*, 1991. Three different preparations of virus from pHXB2-MCS were used in these experiments, namely: MCS, MCS2 and MCS10 (figure 3.8).

Two growth experiments were conducted using two different viral inocula, of 200pg and 20pg per 10^6 cells, respectively, as determined by Sephacryl S1000 chromatography (McKeating *et al*, 1991). Virus production was monitored over a 10 day period of infection in C8166 cells by monitoring for the production of p24 in the culture supernatant by ELISA (McKeating *et al*, 1991).

As can be seen from figures 3.8a and 3.8b, the growth curves are virtually identical between the three pHXB2-MCS preparations and the parental virus pHXB2-SPT. This confirms the infectivity of the new construct and that the nucleotide changes made to create the cassette-vector have no detectable effect on the kinetics of virus production.

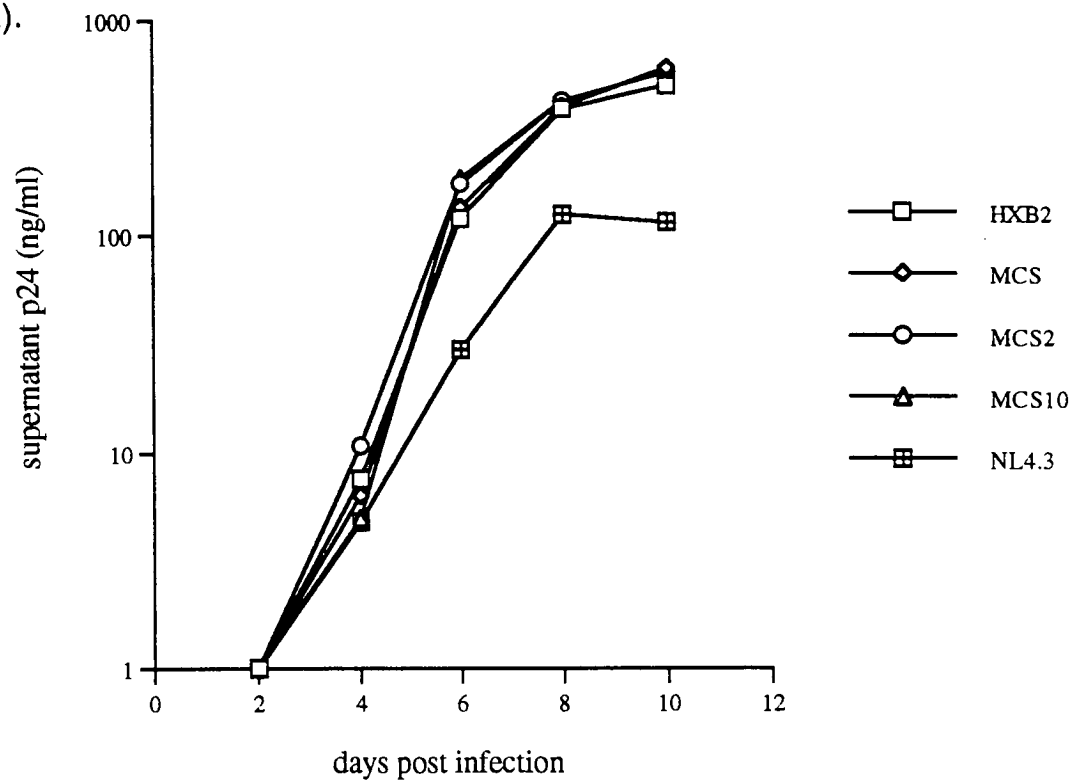
Figure 3.8 The Time Course of p24 Production Following the Infection of C8166 Cells with Virus Produced from Clones pHXB2-SPT, pHXB2-MCS and pNL4-3

a) The time course following inoculation with 200pg virion p24, and
b) inoculation with 20pg virion p24 (*).

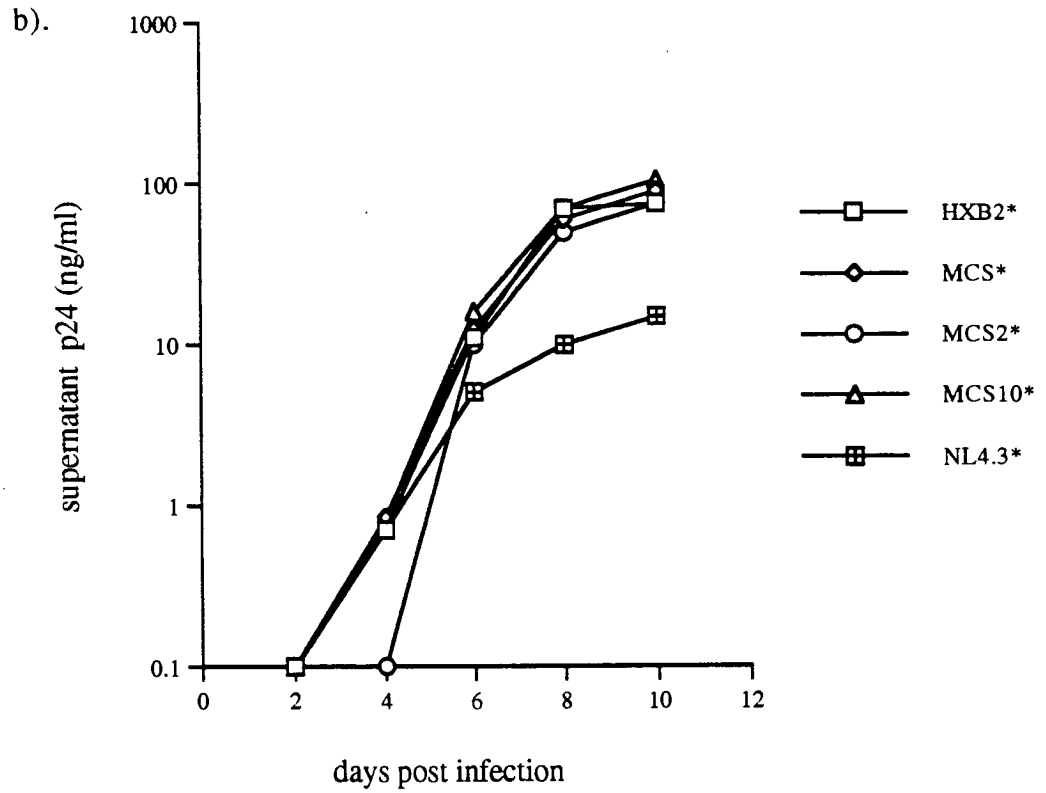
MCS, MC2 and MCS10 refer to the virus derived from three separate viral stocks produced from the transfection of pHXB2-MCS into C8166 cells.
HXB2 refers to virus derived from pHXB2-SPT.

opposite page 118

a).


1000
100
10
1
supernatant p24 (ng/ml)
0
2
4
6
8
10
12
days post infection
HXB2
MCS
MCS2
MCS10
NL4.3

b).


1000
100
10
1
0.1
supernatant p24 (ng/ml)
0
2
4
6
8
10
12
days post infection
HXB2*
MCS*
MCS2*
MCS10*
NL4.3*

Chapter 4

Cloning of *env* Genes

4.1. Introduction

In this chapter, the cloning of *env* gene sequences found over the course of infection in a single HIV-1-infected individual is described. I have chosen to clone the *env* gene sequences from patient 82, a haemophiliac infected in 1984 (Ludlam *et al*, 1985), because two previous studies have characterised the sequence evolution of the *env* gene within this patient, over a 7-year period of infection (Simmonds *et al*, 1991; Holmes *et al*, 1992). This characterisation has allowed the major *env* gene sequences to be identified and cloned in this study.

I have used five peripheral blood mononuclear cell (PBMC) DNA samples to obtain the *env* gene sequences. Although our previous knowledge of the turnover of viral variants was obtained from variants found in the plasma, we observed that most sequences seen in the plasma subsequently appear in PBMC DNA (Simmonds *et al*, 1991). It is not currently feasible to amplify full-length *env* gene sequences from single RNA genomes, but I was able to amplify all major sequences from PBMC DNA, and these were generally intact coding sequences.

Abundance of HIV-1 Target Sequences in Peripheral Blood

HIV sequences are generally in very low abundance in peripheral blood (Brinchman *et al*, 1991; Simmonds *et al*, 1990a; Ho *et al*, 1989). During the asymptomatic phase of infection, proviral DNA copy numbers in peripheral blood generally range from 1 copy per 500 PBMCs to as low as 1 copy 200,000 PBMCs. The proviral load is higher at primary infection and increases to a similarly high level during the symptomatic phases of ARC and AIDS (Bieniasz *et al*, 1993, Haigwood *et al*, 1993, Clark *et al*, 1991; Daar *et al*, 1991). These levels are generally not sufficient to enable cloning of the viral DNA without a method of amplification of the proviral copy number. There are a few exceptions, for example Li *et al*, 1992, were able to clone HIV-1 proviral sequences directly from brain tissue of a post-mortem sample from an AIDS patient, where proviral load was exceptionally high.

In consequence, many HIV-1 clones have been derived from DNA prepared from viral isolates and not directly from patient material (Shaw *et al*, 1984; Levy *et al*,

1986; Adachi *et al*, 1986; Saiki *et al*, 1988; Cheng-Mayer *et al*, 1989; Groenink *et al*, 1991; Gao *et al*, 1996). Virus isolates are produced by culturing the patient PBMC sample with uninfected PBMCs or with a CD4+ immortalised cell line. Virus present in the infected PBMCs should be stimulated to replicate, increasing the viral copy number to sufficient levels for cloning through conventional means of screening genomic libraries constructed from cellular genomic DNA obtained from the viral isolate.

This method has the advantage of increasing the proportion of infectious proviral sequences (Sabino *et al*, 1994). However, culturing results in a marked reduction in the number of variants that were present in the original material (Meyerhans *et al*, 1989, Kusumi *et al*, 1992, Sabino *et al*, 1994). If there is genetic variation in the patient sample which confers differential replicative abilities on the viruses in culture, the subset with the highest replication rate will predominate. For the same reasons, variants that grow to low levels in culture will be under-represented in an isolate.

Therefore, to obtain a representative sample of the spectrum of variants present *in vivo*, the variants must be obtained without prior culturing. This can be achieved by using the polymerase chain reaction (Saiki *et al*, 1985, Mullis and Faloona, 1987) amplifying proviral sequences directly from the PBMC DNA (Ou *et al*, 1988, Meyerhans *et al*, 1989, Simmonds *et al*, 1990a). The polymerase chain reaction is an *in vitro* method for the enzymatic synthesis of specific DNA sequences in which two oligonucleotide primers that flank the region of interest in the target DNA are used to initiate DNA synthesis. A repetitive series of cycles involving DNA template denaturation, primer and template annealing, and the extension of the annealed primers by DNA polymerase, results in the exponential accumulation of copies of the target sequence whose termini are defined by the 5' ends of the primers. The PCR is sufficiently sensitive to allow the direct amplification of viral sequences from uncultured patient material.

Although the spectrum of variants obtained by PCR-amplification of sequences directly from uncultured material should not be biased towards a specific subset of variants, there are problems concerning the accuracy of the amplification. There are

two main sources of sequence error associated with the PCR: firstly, the mis-incorporation of nucleotides during synthesis of the target sequences (Ennis *et al*, 1990); and secondly, the formation of artefactual hybrid sequences by PCR-recombination between different target sequences amplified in the same reaction (Meyerhans *et al*, 1990).

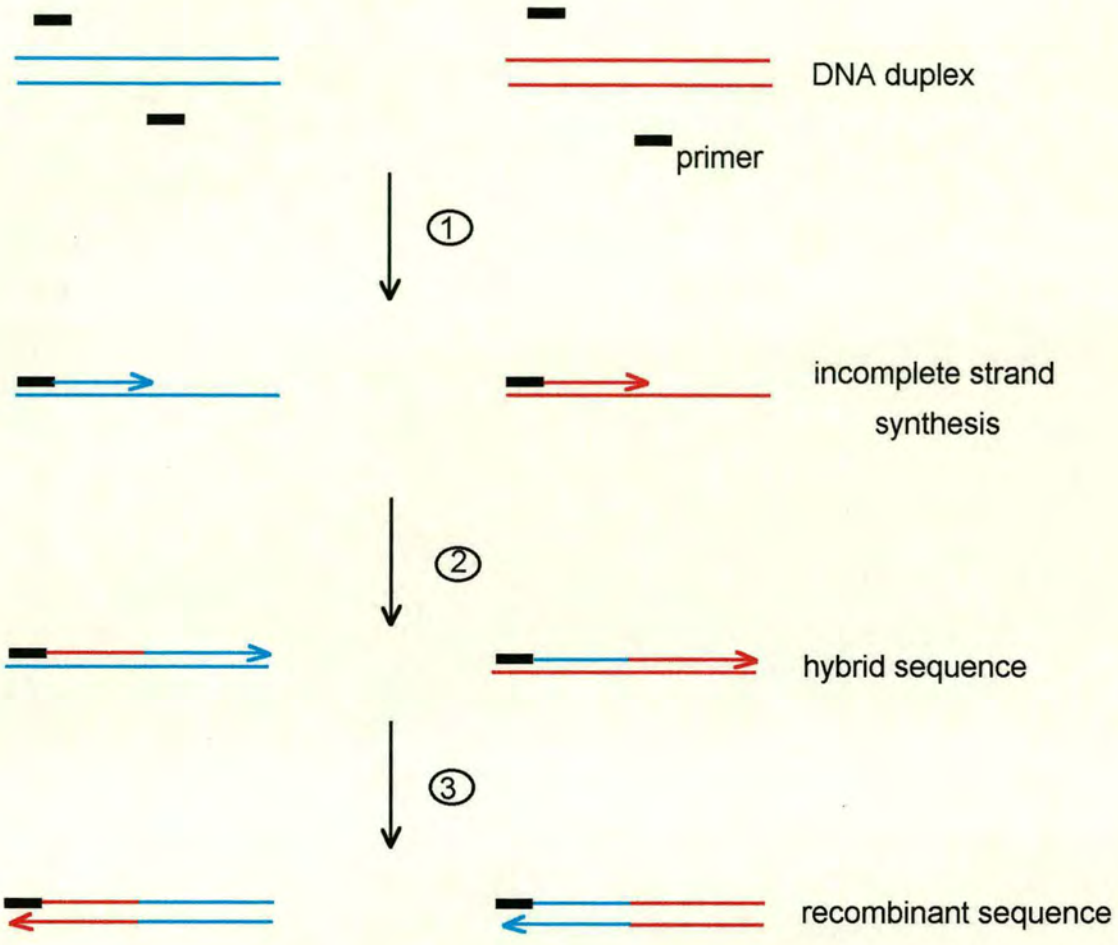
Errors Introduced by PCR Amplification

Misincorporation The heat-stable DNA polymerase, *Taq* polymerase, is the most commonly used DNA polymerase for PCR. *Taq* polymerase does not contain a 3'->5'exonuclease, or 'proof-reading', ability, and therefore has a relatively high mis-incorporation rate. The misincorporation rate has been measured as 1 in 400 (Saiki *et al*, 1988) to 1 in 1800 (Ennis *et al*, 1990), depending on the individual reaction conditions. A means of reducing, or controlling for, this high error-rate is important when the amplified sequence is to be cloned and expressed, as is the case in this study, as these errors may lead to an aberrant or defective protein being expressed.

***In Vitro* Recombination** In addition, the amplification of a heterogeneous population of sequences, such as the population of viral genomes found *in vivo*, (Saag *et al*, 1988, Alizon *et al*, 1986, Hahn *et al*, 1986, Balfe *et al*, 1990), can result in the production of artefactual recombinant molecules, due to recombination between different sequences occurring during the PCR (Meyerhans *et al*, 1990, Ennis *et al*, 1990). These false-recombinant sequences are thought to arise from the premature termination of *Taq* polymerase in one cycle, and the hybridisation of this unfinished strand to a heterologous strand in the next cycle. Following the subsequent extension and completion of the strand a hybrid sequence will be produced. This process is illustrated in figure 4.1.

A high rate of *in vitro* recombination was described by Meyerhans *et al*, 1990, with 1 in 5 sequences being an artificial recombinant after 25 cycles, amplifying a 300bp stretch of DNA from reactions containing two different cloned variants of the HIV-1 *tat* gene. It is to be expected that the frequency of false-recombinants will increase with an increase in length of the amplified fragment, as there will be an

Figure 4.1. A Schematic Representation of the Formation of Recombinant Sequences during PCR



Two variants (blue and red) of a given genomic fragment are shown.

1. Following a round of PCR (denaturation, annealing and elongation), some incomplete strand synthesis occurs.

2. In a subsequent round of PCR, annealing of incomplete strands to heterologous templates, followed by elongation, will generate hybrid sequences.

3. A further round of PCR will generate recombinant sequences.

increase in the probability of incomplete strand synthesis. It is therefore necessary to control for, or reduce, the generation of artefactual recombinant sequences in order to amplify sequences accurately by PCR.

In the amplification of *env* gene sequences in this study, two modifications to the standard PCR were carried out to reduce the errors inherent in the PCR-amplification of sequences. Firstly, each amplification was carried out from a single target molecule. Amplification from a single molecule avoids the possibility of producing artefactual recombinants by recombination *in vitro*. Recombination may occur between strands, but template-switching, in this case, will have no effect on the end sequence, as all templates are derived from the same initial molecule.

Secondly, an alternative thermostable DNA polymerase, *Pfu* polymerase, was used which has a proof-reading ability, and consequently, a misincorporation-rate, measured as 12-fold lower than *Taq* polymerase (Lundberg *et al*, 1991, Mathur *et al*, 1992).

In addition, by starting from a single target sequence, it has been possible to calculate the number of errors by sequencing of two or more clones derived from the same amplification and examining these for sequence differences.

4.2. Results

4.2.1. Single-Molecule Amplification of *env* Gene Sequences

In this study the limiting-dilution, nested-PCR of Simmonds *et al*, 1990a was used to amplify from single molecules, the 1.7kb (cassette-1) and the 2.5kb (cassette-2) *env* gene fragments.

In this method, the PBMC DNA is diluted before amplification and distributed into a number (usually 30 to 50) of replicate reaction tubes. At a specific 'end-point' dilution (dependent on the proviral copy number and the concentration of the DNA sample), only a small proportion of the reactions will contain a target sequence and so yield amplified DNA. By assuming a random distribution of proviral sequences in the reaction tubes, it is possible to calculate, by reference to the Poisson distribution, the probability that each PCR-positive amplification was initiated from a single molecule. In this laboratory, we consider the DNA sample to be sufficiently dilute, ie. at its end-point dilution, when the frequency of PCR-positive amplifications is 0.2 or less. At this value, there is a 90% probability that each reaction initiated from a single molecule (Simmonds *et al*, 1990a).

A critical factor in the accuracy of the limiting-dilution method, is the ability to detect the DNA following amplification from a single target molecule. In this laboratory, we employ a nested-PCR, in which two rounds of cycling is carried out with the primers for the second round located within the first round pair. This second round increases both the yield and specificity of the product. It was demonstrated by Simmonds *et al*, 1990a, that a nested-PCR, amplifying a 278bp fragment of the *env* gene across the V1-V2 hypervariable regions, is capable of amplifying to detectable levels, visible by agarose gel-electrophoresis, from single target molecules.

To obtain an amplification of the 1.7kb and 2.5kb fragments of the *env* gene from single molecules, the PBMC DNA samples were first titrated to the end-point dilution, using the V1-V2 nested-PCR described above, previously demonstrated to

be sufficiently sensitive to amplify from single target molecules. Once the end-point dilution was reached, this dilution was used in a nested-PCR with the primers specified to amplify the 1.7kb or 2.5kb fragment. By altering the reaction conditions, that is, conducting the annealing step at 45^0C and allowing DNA extension to proceed for 4 mins, two rounds of 25 cycles were capable of amplifying the 1.7kb fragment to detectable levels, visible by agarose gel electrophoresis. The 2.5kb fragment could not, however, be amplified consistently from single molecules. Therefore, the majority of *env* gene sequences examined in this study were of the 1.7kb fragment.

Table 4.1. shows the end-point dilutions calculated for each of the 5 PBMC DNA samples used in this study, with the frequency of PCR-positives obtained from the nested-PCR amplifying with the V1-V2 primers. This is compared to the frequency of PCR-positives generated by amplifying the DNA at its end-point dilution with the 1.7kb PCR. For each DNA sample, the number of positives generated in both PCRs was similar, demonstrating that the 1.7kb PCR can efficiently amplify from single molecules. The frequency of PCR-positive reactions was in all cases 0.2 or less, which, as described above, gives a probability that 90% of these were initiated from a single molecule.

Table 41. A Comparison of the Efficiency of Amplification of the V1-V2 (278bp) Region and the 1.7kb Fragments of the *env* Gene

PBMC sample date	PBMC DNA sample number	end-point dilution[a]	no. of PCR-positives/total no. of reactions	
			V1-V2[b]	1.7kb[c]
1987	#108	1/15	7/30 (0.23)	3/30 (0.01) 5/28 (0.18)
1988	#12	1/50	1/5 (0.20)	2/25 (0.08)
1989A	#82	1/500	2/10 (0.20) 5/30 (0.17)	[b]3/25 (0.12) 3/25 (0.12)
1989B	#123	1/200	1/5 (0.20) 0/20 (0.00)	2/25 (0.08) 3/25 (0.12) 3/25 (0.12) 1/25 (0.04)
1990	#139	1/50	1/5 (0.20) 1/25 (0.04)	3/25 (0.12) 4/25 (0.16)

The figures in brackets refer to the frequency of positive reactions.

[a] at the end-point dilution each PCR contains 1ul of the DNA dilution given for each PBMC DNA sample

[b] Taq amplified

[c] Pfu amplified

Note on the Nomenclature of the *env* Gene Clones

The remaining sections of this chapter describe the cloning of the PCR-amplified *env* gene variants and their subsequent incorporation into pHXB2-MCS.

A complex naming-scheme was adopted to identify each amplified *env* sequence, and the clones derived from it. This naming-scheme has been maintained in this chapter to facilitate the descriptions of the cloning steps, and to allow identification of each clone in future work.

The complex naming-scheme is as follows:
a) each amplified *env* gene sequence (ie. each PCR-positive) is referred to by two numbers (w.x). The first number (w) identifies the PBMC DNA sample from which the *env* sequence was amplified; the second number (x) distinguishes each individually-amplified *env* gene sequence from that DNA sample. For example, in table 4.2, '082.03' refers to *env* sequence number 03, amplified from PBMC DNA sample #82. '082.06' refers to *env* sequence number 06, from PBMC DNA sample #82.

b) *env* gene subclones are referred to by a prefix 'p', followed by three numbers (p.w.x.y). The first number (w) gives the PBMC sample number; the second number (x) the number of the amplified *env* sequence; the third (y) distinguishes each clone obtained from the amplified *env* sequence. For example, p082.03.03 refers to clone number 03, derived from the cloning of *env* sequence number 03, amplified from PBMC DNA sample #82. p082.03.05 refers to clone number 05, derived from the same amplified *env* sequence. As such, these two clones should be identical if there has been no PCR errors.

c) recombinant proviral clones - generated following the substitution of *env* gene clones into pHXB2-MCSΔ*env* - are denoted by a prefix 'H' followed by four numbers (H.w.x.y.z). The first number (w) gives the PBMC sample number; the second (x), the number of the amplified *env* sequence; the third (y) distinguishes each

clone obtained from the amplified *env* sequence; the fourth number (z) distinguishing each recombinant proviral clone recovered from the transfer of the cloned *env* sequence into the cassette-vector. For example, H.082.03.05.01 refers to recombinant proviral clone number 01, obtained from the incorporation of the *env* gene from subclone p082.03.05 into the cassette-vector pHXB2-MCSΔ*env*.

I have also adopted a more simplified nomenclature whereby each *env* sequence is given a letter (A to S) followed by a number (1 to 3) to show its length (1=1.7kb; 2=2.5kb; 3=1.4kb).

All *env* clones and full-length proviral clones generated in this study are summarised in table 4.8, at the end of this chapter.

4.2.2. Cloning and Characterisation of 1.7kb *env* Gene Sequences

a) *Taq* Polymerase-Derived Clones

The preliminary amplification of 1.7kb *env* gene sequences was carried out with *Taq* polymerase, from patient PBMC DNA sample #82 (1989A). At the end-point dilution, three out of a total of twenty-five amplification reactions were positive, each yielding a single band of DNA at 1.7kb (table 4.1). The details of experimental conditions are given in chapter 2, section.

The products of these three positive PCR reactions were digested with *Bst*EII and *Xba*I and cloned in the plasmid vector pNBXXΔ*env* (chapter 3, section 3.6), by replacement of the deletion-containing 1.3kb *Bst*EII-*Xba*I vector *env* fragment. Two, eleven and four clones were recovered from the three amplified *env* sequences, respectively (table 4.2).

Assessment of *Taq* Amplification Errors

The number of errors arising from the PCR-amplification was assessed by sequencing two clones from each of the three amplified *env* gene variants. Each clone was sequenced across a 150bp stretch of the third hypervariable (V3) region and across a 200bp region encompassing the fourth and fifth hypervariable regions and the fourth conserved region C4 (V4-C4-V5). Errors were detected by comparing the sequence of clones derived from the same *env* variant.

A total of four nucleotide differences were found between all comparisons made between clones derived from the same *env* variant. These differences are summarised in table 4.3; the actual nucleotide sequences are given in appendix D.

No nucleotide differences were found between clones p082.03.03 and p082.03.05 in either the V3 or V4-C4-V5 region. Three nucleotide differences were found between clones p082.06.11 (L.1) and p082.06.12 (M.1). Two of the these differences - C and

Table 4.2 *Taq*-Amplified 1.7kb *env* Gene Sequences

PBMC sample	*env* variant	total number clones obtained	clones examined further
#82 (1989A)	082.03	2	p082.03.03, -05
	082.06	11	p082.06.11, -12
	082.10	4	p082.10.13, -22

Table 4.3 Nucleotide Differences Between *Taq*-Amplified Clones Derived from the same PCR Amplification.

env variant	clones	region	V3		V4-C4-V5	
		nucleotide position	7184	7220	7531	7561
082.06 (L.1)	p082.06.11		A (Q)	C (I)	-	A (I)
(M.1)	p082.06.12		T (R)	T (I)	-	G (I)
082.10 (N.1)	p082.10.13		-	-	T (I)	-
(W.1)	p082.10.22		-	-	A (N)	-

The nucleotide position of the observed differences are given according to HXB2R. The letters in brackets refer to the amino acid residue at each site. Where '-' is marked no differences were found.

T (position 7220) and A and G (at 7561) - are synonymous codon changes. The third nucleotide difference - A or T (at 7184) - results in an amino acid difference - aspartate (Q) or arginine (R) in the deduced sequence. One nucleotide difference was observed between clones p082.10.13 (N.1) and p082.10.22 (W.1) - T to A (at position 7531) which results in a codon change of I to asparagine (N).

There are two possible reasons for the sequence discrepancies between clones derived from the same PCR-amplification. Either, despite amplifying at the endpoint dilution, there are two variants in the original reaction, or alternatively, the differences are the result of nucleotide misincorporation by *Taq* polymerase.

Two of the four nucleotide differences result in amino acid residues that have not been observed before in sequences in the data sets from patient 82 (Simmons *et al*, 1991; Holmes *et al*, 1992; Zhang, 1993). Firstly, an A at position 7184 (L.1 or p082.06.11) - which results in an arginine (R) amino acid residue at position 25 of the V3 loop sequence is not seen in 87 V3 sequences derived from this patient (Zhang, 1993). Secondly, the T at 7531 (N.1 or p082.10.13) which results in an asparagine (N) residue in the conserved region C4 is not found the data set of 114 V4-C4-V5 sequences obtained by Zhang, 1993. The absence of these nucleotides in the patient 82-data set of Zhang, 1993, provides evidence that suggests that they are the product of mis-incorporation by *Taq*-polymerase.

Assuming each of the four nucleotide differences in sequence between clones amplified from the same reaction are *Taq*-induced errors, then four nucleotides mis-incorporations have occurred over a length of 2200 nucleotides examined (or a mean of 1 error per 525 nucleotides). If just two of these (the nucleotides resulting in an unusual amino acid) are *Taq*-induced errors, then the mean number of errors is 1 per 1050 nucleotides amplified and cloned.

These values are in agreement with other published work. For example Saiki *et al*, 1988 found 1 error per 400 nucleotides, and Ennis *et al*, 1990 found 1 error per 1800 nucleotides.

The high number of errors produced following amplification with *Taq* polymerase means that *Taq* is unsuitable for amplifying sequence variants for their subsequent expression, as on average no cloned variant will represent the original sequence found *in vivo*. All subsequent PCR amplifications were carried out using *Pfu*-polymerase, from which I detected less than 1 error per 16kb of amplified and cloned sequence, as will be described below.

4.2.2. Cloning and Characterisation of 1.7kb *env* Gene Sequences (Continued)

b) *Pfu* Polymerase-Derived Clones

A total of 29 single molecules were isolated by limiting-dilution from 5 PBMC samples from p82, and amplified over the 1.7kb region of the *env* gene with *Pfu* polymerase (table 4.1). To obtain representative examples of the major V3 forms identified at each of the 7 time points studied in an earlier longitudinal study of V3-variability in this patient (Holmes *et al*, 1992) 13 of the 29 isolated molecules were cloned (table 4.4). (The characterisation of these clones according to the sequence of their V3, V4 and V5 hypervariable regions is given in the chapter 5).

Each amplified proviral variant was digested with *Bst*EII and *Xba*I and cloned into the plasmid vector pNBXXΔ*env*, replacing the deletion-containing 1.3kb *Bst*EII/*Xba*I vector *env* fragment. Between one and twelve clones were recovered from each proviral variant, according to individual cloning efficiencies (table 4.4).

Table 4.4 1.7kb Single-Molecule Derived, *Pfu*-Amplified *env* Gene Variants

This table shows the 13 *env* gene variants amplified with *Pfu*-polymerase and the number of clones obtained from each.

opposite page 134

year	PBMC sample	*env* variant	total number clones obtained	clones examined further
1987	#108	108.18 (A.1)	2	p108.18.27, .28
		108.11 (B.1)	2	p108.11.03, .07
		108.10 (C.1)	3	p108.10.04, .06, .07
		108.15 (D.1)	2	p108.15.01, .04
		108.17 (E.1)	2	p108.17.32, .33
		108.19 (F.1)	3	p108.19.04, .05, .06
1988	#12	012.08 (T.1)	1	p012.08.10
		012.16 (U.1)	3	p112.16.03, .04, .14
1989A	#82	082.24 (G.1)	12	p082.24.01, .04
1989B	#123	123.14 (H.1)	3	p123.14.09, .10, .11
1990	#139	139.05 (I.1)	6	p139.05.01, .02, .16,.20
		139.06 (V.1)	6	p139.06.08, .10, .11
		139.6b (J.1)	6	p139.6b.09, .12

Assessment of *Pfu* Amplification Errors

To determine the number of errors resulting from amplification, and to ensure that clones, where expected, were derived from the same proviral molecule, multiple clones derived from the amplification reaction were sequenced across at least four out of the five hypervariable regions of the *env* gene, and also across a 200bp stretch within the gp41 coding region. Table 4.5 summarises the regions sequenced from each *env* clone. (The actual nucleotide sequences are given in appendix D).

The number of errors was calculated by making sequence comparisons between clones derived from the same proviral variant. However, no sequence differences were detected between clones derived from the same proviral variant, despite there being sequence variation between the individual proviral molecules. The length compared covered a total of 16.32kb of cloned sequence (table 4.5). From this we can conclude that *Pfu* polymerase has a very low error rate and should be suitable to amplify 1.7kb *env* gene sequences without the introduction of errors.

Table 4.5 Summary of the Numbers of Clones from Each *env* Variant Sequenced in One or More of Five Regions of the *env* Gene

opposite page 136

	number of clones sequenced				
	region	V1/V2	V3	V4/C4	V5	gp41
env variant	length (bp)	180	200	200	90	200
108.18 (A.1)		2	2	2	1	2
108.11 (B.1)		2	2	2	2	1
108.10 (C.1)		2	2	3	1	0
108.15 (D.1)		2	1	1	0	2
108.17 (E.1)		1	1	2	1	2
108.19 (F.1)		1	2	2	1	1
012.08 (T.1)		1	1	1	1	1
012.16 (U.1)		2	2	3	1	3
082.24 (G.1)		1	1	1	1	1
123.14 (H.1)		2	3	3	2	1
139.05 (I.1)		1	4	4	1	4
139.06 (V.1)		0	3	5	1	3
139.6b (J.1)		0	2	4	1	2
total number of clones		18	27	34	14	23
total length sequenced (kb)		3.24	5.40	6.80	1.26	4.60
total number of clones compared		12	22	29	4	18
total length compared (kb)		2.16	2.60	5.8	.36	3.6

Total length sequenced = 21.3kb

Total length compared (between clones derived from the same amplification) = 16.32kb

4.2.2. Cloning and Characterisation of 1.7kb *env* Gene Sequences (Continued)

c) Characterisation of *Pfu*-Amplified Clones by Restriction-Mapping

The thirteen *Pfu*-amplified *env* variants were characterised by restriction-mapping in the pNBXX-background. Several restriction enzymes were used (*Bam*HI, *Bsm*I, *Hin*dIII, *Kpn*I, *Mlu*I, *Nde*I, *Nhe*I, *Not*I, *Sal*I, *Xba*I, *Xho*I, *Eco*RI, and *Ssp*I). Multiple clones derived from the same amplification were restriction-mapped. With each variant *env* gene, clones from the same amplification produced identical restriction patterns, further demonstrating that each amplification was initiated from a single proviral sequence (data not shown).

Eleven clones, as expected, showed the correct sized insert fragment of 1.7kb. Two (T.1 and U.1), however, contained anomalous bands which had not been detected earlier. In consequence, no futher analysis was carried out with these two clones.

Four restriction enzymes: *Bsm*I, *Hin*dIII, *Kpn*I and *Ssp*I, were found to show different cleavage patterns (restriction-site polymorphism) between the panel of clones (table 4.6). *Bsm*I shows two polymorphic sites: one at nucleotide 306 (where co-ordinate 1 is the start of the cassette at the *Bst*EII site), and one at 1731. *Hin*dIII shows a single polymorphic site at 237. *Kpn*I shows two polymorphic sites: one at 22 and approximately 1000, although the site at 1000 was found in only one variant: I.1. *Ssp*I shows two polymorphic sites: one at 1251, which was present in nine out of the 13 clones, and the other polymorphic site at approximately 1700, although this site is only seen in variant I.1.

In total, the four restriction enzymes produced nine different restriction patterns for the set of eleven *env* variants (table 4.6). Six clones (C.1, D.1, F.1, I.1, V.1 and J.1) and pNBXX, show a unique restriction pattern (patterns 3, 4, 5, 7, 8, 9 and 1, respectively). Variants G.1 and H.1 show the same restriction pattern (pattern 6); and restriction pattern 2 is shared between three clones A.1, B.1 and E.1.

Table 4.6 Restriction-Enzyme Recognition Sites

[1] co-ordinate 1 is the *Bst*EII site at the 5'end of the cassette

+ presence of site, - absence of site

opposite page 138

	*Bsm*I		*Hin*dIII	*Kpn*I		*Ssp*I		
pattern number	306[1]	1731	237	22	1000	1251	1700	clones
1	-	+	-	+	-	+	-	pNBXX
2	+	+	-	+	-	+	-	A.1, B.1, E.1
3	-	+	-	+	-	+	-	C.1
4	+	+	-	-	-	+	-	D.1
5	+	+	+	+	-	+	-	F.1
6	+	+	+	+	-	+	-	G.1, H.1
7	-	+	-	-	+	-	+	I.1
8	+	+	+	+	-	-	-	V.1
9	-	-	+	+	-	+	-	J.1

4.2.2. Cloning and Characterisation of 1.7kb *env* Gene Sequences (Continued)

d) Examination of Coding Potential by *In Vitro* Transcription/Translation

The technique of *in vitro* transcription followed by *in vitro* translation was applied to screen the panel of *env* gene clones for an intact reading frame.

The design of the plasmid vector, pNBXXΔ*env*, into which the *env* variants were cloned, is such that the entire *env* gene can be transcribed using T7 RNA polymerase from the T7 promoter, which is located upstream of the *env* gene (chapter 3, section 3.6). The first start codon of the transcripts is that of the *env* gene, and as such, can be used to direct *in vitro* translation.

Despite numerous attempts, this technique was successful on only one occasion. The succesful experiment was conducted with three clones: p123.14.09, p123.14.10 and p123.14.11 - which are clones of the same *env* variant H.1 or 123.14 (table 4.4) - and the control plasmid pNBXX. Each clone was linearised by digestion with *Xho*I, and transcribed with T7 RNA polymerase. The resultant transcripts were added to a wheat-germ translation system, the details of which can be found in chapter 2, section 2.10.

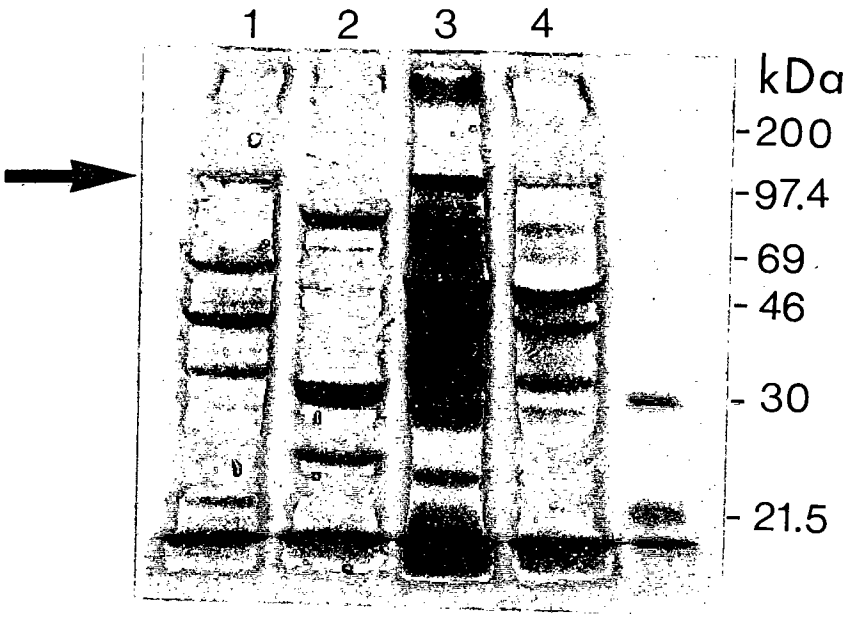
An autoradiograph of the translation products is shown in figure 4.3. The control clone pNBXX, and clones p123.14.10 and p123.14.11, produced a translation product of approximately 100kDa (arrow), the size of non-glycosylated gp160. Clone p123.14.09, however, produced a shorter product of approximately 80kDa, which may be the result of a truncated *env* gene. If so, this mutation was not picked up during sequencing this clone (section 4.2.2, above). The autoradiograph also shows the presence of many shorter protein products, which may have resulted from one of the following three processes: proteolysis, internal ribosomal initiation, or premature termination of translation.

Figure 4.3. Autoradiograph Showing the Result of *In Vitro* Transcription/Translation of *env* Clones

Lane 1: pNBXX; lane 2: p123.14.09; lane 3: p123.14.10; lane 4: p123.14.11

The arrow indicates the presence of a protein of approximately 100kDa.

opposite page 140


1
2
3
4
kDa
200
97.4
69
46
30
21.5

As this technique did not produce a clear predominant band of the expected size, and was difficult to repeat, I proceeded with the next cloning step of substituting the *env* variants into pHXB2-MCSΔ*env* and determining their infectivity following transfection (as will be described in chapter 6).

4.2.3. Incorporation of 1.7kb *env* Genes into pHXB2-MCSΔenv

a) With the *Pfu*-Amplified Clones

Ten of the thirteen *Pfu*-derived 1.7kb *env* genes (A.1 to J.1) were transferred to the cassette-vector pHXB2-MCSΔ*env* (a description of which was given in chapter 3, section 3.6), to construct ten full-length recombinant proviral clones (table 4.8 (1) below).

The ten variants were chosen on the basis of their V3 loop sequences to include the major V3 sequence types detected in the plasma over the course of infection within patient 82. The sequence of the clones will be discussed further in the next chapter.

Each clone was digested with *Bst*EII and *Xba*I, the 1.7kb *env* fragment was isolated and purified by agarose-gel electrophoresis. The purified fragments were then ligated into the vector pHXB2-MCSΔ*env*, replacing the shorter 1.3kb *Bst*EII-*Xba*I vector fragment.

The identity of the *env* variant in the pHXB2-MCS background was confirmed by restriction-mapping with the same panel of enzymes that were used to generate the maps in the pNBXX background (table 4.6, above).

b) With the *Taq*-Amplified Clones

In addition to the *Pfu*-derived *env* sequences, four *Taq*-amplified variants, K.1, L.1, M.1, and N.1 (table 4.2, above), were used to construct recombinant clones in the pHXB2-MCS background. This was carried out to obtain additional clones that have a V3-loop sequence predictive of an syncytium-inducing phenotype (chapter 5). No

additional clones were available from the *Pfu* amplified clones, therefore, the *Taq* variants were taken despite the high error rate associated with the amplification with *Taq* polymerase.

The four clones were constructed in an identical manner to the *Pfu*-derived clones: first they were digested by *Bst*EII and *Xba*I, the 1.7kb *env* frgament gel-purified, and cloned into pHXB2-MCSΔ*env*, replacing the shorter 1.3kb *Bst*EII/*Xba*I vector fragment (table 4.8 (1)).

c) 'Reconstruction' of pHXB2-MCS

A single preparation of the *Bst*EII-*Xba*I digested and purified vector pHXB2-MCSΔ*env* was used to construct all the 1.7kb recombinant full-length clones described in this study. Following the construction of the above 14 recombinant proviral clones, an aliquot of the *Bst*EII/*Xba*I digested pHXB2-MCSΔ*env* was tested to ensure it was fully-intact and capable of forming an infectious provirus. To do this, the 1.7kb *Bst*EII-*Xba*I fragment of pNBXX was isolated and cloned into pHXB2-MCSΔ*env*, replacing the shorter *Bst*EII-*Xba*I 1.3kb vector fragment - thus 're-creating' pHXB2-MCS. This proviral clone was found to be capable of producing infectious virus on transfection into cell culture, confirming its viability (data not shown).

4.2.4. Substitution of 2.5kb *env* Variants into pHXB2-MCSΔ*env*

In addition to the construction of the fourteen recombinant viruses with the 1.7kb fragment of the *env* gene derived from PBMC-DNA, amplification of cassette-2 (2.5kb) encoding the entire *env* gene, and part of *tat*, *rev* and *nef*, was attempted from single proviral molecules.

Despite the relative efficiency at which the 1.7kb fragment could be amplified, the 2.5kb fragment proved extremely difficult to amplify from PBMC DNA at its end-point dilution. After many attempts, one sequence, P.2 (108.42), was amplified across the 2.5kb region, from PBMC sample #108. The PCR product was digested with *Bst*EII and *Xho*I and cloned into pNBXXΔ*env*, replacing the shorter *Bst*EII-*Xho*I 2.3kb vector fragment. A single clone, p108.42.18, was recovered. This was partially-sequenced and restriction mapped (see below). The 2.5kb *env* variant was then transferred to the cassette-vector pHXB2-MCSΔ*env*, replacing the shorter 2.1kb, vector, *Bst*EII-*Xho*I fragment (table 4.8 (2)).

Three additional clones were also constructed with 2.5kb *env* gene fragments from PBMC DNA sample #82. It was not possible with this PBMC DNA sample to amplify the 2.5kb region from single molecules. Therefore, a bulk nested-PCR amplification of the PBMC DNA was carried out, by amplifying 1ul of PBMC DNA with cassette-2 outer primers, in a single reaction. From the limiting-dilution calculation for this PBMC sample given in table 4.1, above, this PCR reaction will contain approximately 500 proviral molecules. Following the second round of PCR with the *Bst*EII/*Xho*I inner primers, a single DNA band of 2.5kb was produced (data not shown), and was digested with *Bst*EII and *Xho*I and cloned directly into the proviral vector pHXB2-MCSΔ*env*, replacing the shorter 2.1kb *Bst*EII/*Xho*I vector fragment. Three recombinant proviral clones were recovered from the cloning. These were designated Q.2, R.2, and S.2 (table 4.8 (2)).

Restriction-Mapping of 2.5kb *env* Clones

Variants P.2, Q.2, R.2 and S.2 were restriction mapped with *Bsm*I, *Hin*dIII, *Kpn*I and *Ssp*I to characterise each variant. The location of the restriction sites given in table 4.7. The same polymorphic sites were seen for these clones as for the 1.7kb clones, namely, two polymorphic sites for *Bsm*I, one at 306 and one at 1731; one *Hin*dIII polymorphic site at 237; one polymorphic site for *Kpn*I at 22; and one polymorphic site 1239 for *Ssp*I. In addition, there is a polymorphic *Ssp*I site at 2110, absent in the pHXB2-MCS sequence but present in the four 2.5kb PBMC-derived clones. There is also a *Bam*HI site at 2100 in pHXB2-MCS, which is not present in each of the four 2.5kb clones. These sites for *Ssp*I and *Bam*HI map outside the 1.7kb fragment of the clones described above (section 4.4.2).

Only one of the three 2.5kb-clones obtained from PBMC sample #82, R.2, has been sequenced (chapter 5, section 5.3). The restriction patterns produced from these three clones reveals that the three clones clearly consist of two variants, with Q.2 and S.2 producing the same pattern (map 12), which is distinct from the restriction pattern of R.2 (map 13). The patterns also demonstrate that although clones Q.2 and S.2 have not been sequenced, they are very similar to the maps of the other clones, providing evidence that they are indeed amplified from patient 82 DNA, and not the product of contamination.

Table 4.7 Restriction Enzyme Recognition Sites

[1] co-ordinate 1 is the *Bst*EII site at the 5' end of the cassette

+ presence; - absence

opposite page 145

	*Bsm*I		*Hin*dIII	*Kpn*I	*Ssp*I		*Bam*HI	
pattern number	306[1]	1731	237	22	1251	2110	2100	clones
10	-	+	-	+	+	-	+	HXB2
11	-	+	+	+	+	+	-	P.2
12	+	+	+	+	+	+	-	Q.2, S.2
13	+	-	+	-	+	+	-	R.2

4.2.5 Subcloning of 1.4kb *env* Variant A into pHXB2-MCS.*Mlu*I

The proviral cassette-vector pHXB2.MCS.*Mlu*I (chapter 3) was used to construct a single recombinant proviral clone, A.3 (H.108.18.27.*Mlu*I), by exchange of the vector *Bst*EII to *Mlu*I fragment with that from *env* clone A.1 or p108.18.27 (table 4.8 (3)). The *Bst*EII to *Mlu*I fragment (from 5875 to 7279 HXB2R) is a 1.4kb fragment of the *env* gene and includes that part encoding for gp120 only (chapter 3).

The 1.4kb fragment was amplified from 10ng of p108.18.27 (A.1) DNA, in a single round of PCR with primers specified at the *Bst*EII and *Mlu*I sites, using *Pfu* polymerase. The amplification product was digested with *Bst*EII and *Mlu*I and cloned into pHXB2-MCS.*Mlu*I, replacing the *Bst*EII/*Mlu*I vector fragment. The identity of the clone was confirmed by restriction digest with enzymes *Bsm*I, *Kpn*I, *Hin*dIII and *Ssp*I.

4.2.6 Substitution of 1.7kb and 2.5kb *env* fragments from HIV-1SF2$_{MC}$ into pHXB2-MCSΔ*env*

Two recombinant proviral clones were generated by substituting the 1.7kb, and the 2.5kb fragment of pHXB2-MCSΔ*env* with the equivalent regions amplified from the T-cell-line tropic infectious molecular clone SF2$_{MC}$ (Cheng-Mayer *et al*, 1989).

To construct the two recombinant viruses, 50 ng of the molecular clone SF2$_{MC}$ were used in a PCR with *Pfu* polymerase, with either the *Bst*EII/*Xba*I primers for the 1.7kb fragment, or the *Bst*EII/*Xho*I primers for the 2.5kb fragment. The PCR products were digested with the appropriate combination of restriction enzymes and cloned into pHXB2-MCSΔ*env*, replacing the vector fragment. The identity of the recombinant proviral clones recovered from each of the two cloning experiments were confirmed by restriction-enzyme analysis (data not shown). The clones were designated HXSF2.1 and HXSF2.2 (table 4.8(4)).

Table 4.8 Summary of the Construction of the HIV-1 Recombinant Proviral Clones

n.d. not done: *env* gene variant was not incorporated into a proviral clone

opposite page 147

(1) Recombinant for Cassette-1 - the 1.7kb fragment of the *env* gene

- With *Pfu*-Amplified *env* Sequences

PBMC sample date	*env* variant	*env* subclone	recombinant proviral clone	name
1987	108.18	p108.18.27	H.108.18.27.02	A.1
	108.11	p108.11.07	H.108.11.07.10	B.1
	108.10	p108.10.07	H.108.10.07.02	C.1
	108.15	p108.15.04	H.108.15.03.44	D.1
	108.17	p108.17.03	H.108.17.33.03	E.1
	108.19	p108.19.04	H.108.19.04.05	F.1
1988	012.08	-	n.d.	(T.1)
	012.16	-	n.d.	(U.1)
1989A	082.24	p082.24.01	H.082.24.01.36	G.1
1989B	123.14	p123.14.10	H.123.14.10.02	H.1
1990	139.05	p139.05.20	H.139.05.20.05	I.1
	139.06	-	n.d.	(V.1)
	139.6b	p139.6b.09	H.139.6b.09.09	J.1

- With *Taq*-Amplified *Env* Sequences

PBMC sample date	*env* variant	*env* subclone	recombinant proviral clone	name
1989A	082.03	p082.03.03	H.082.03.03.10	K.1
	082.06.11	p082.06.11	H.082.06.11.09	L.1
	082.06.12	p082.06.12	H.082.06.12.04	M.1
	082.10	p082.10.13	H.082.10.13.23	N.1

(2) Recombinant for Cassette-2 - the 2.5kb fragment of the *Env* gene

PBMC sample date	*env* variant	*env* subclone	recombinant proviral clone	name
1987	108.42	p108.42.18	H.108.42.18.35	P.2
1989A	82.13 82.14 82.17	n.d. n.d. n.d.	H.082.13 H.082.14 H.082.17	Q.2 R.2 S.2

(3) Recombinant for Cassette-3 - the 1.4kb fragment of the *Env* gene

PBMC sample date	*env* variant	*env* subclone	recombinant proviral clone	name
1987	108.18	p108.18.27	H.108.18.27.*Mlu*I	A.3

(4) Recombinant for Cassette-1 (1.7kb) and Cassette-2 (2.4kb) - with *Env* gene fragments from $SF2_{MC}$

env gene variant	recombinant proviral clone
$SF2_{MC}$ (1.7kb) $SF2_{MC}$ (2.4kb)	HXSF2.1 HXSF2.2

Chapter 5

Sequence Analysis of the *env* Genes

5.1. Introduction

Two previous studies have characterised in detail the evolution of the virus population within patient 82. In the first study, Simmonds *et al*, 1991, have described the evolution of the V4 and V5 hypervariable regions of the *env* gene of the PBMC-, and plasma-associated viral populations, over a 6 year period of infection. In the second study, Holmes *et al*, 1992, have described the sequence evolution of the V3 hypervariable region, of virus found in the plasma from the time of seroconversion, and in 7 subsequent years of infection.

The results show firstly, that there is considerable variation within each sample, at both the nucleotide and amino acid levels, and a rapid turnover of variants with time. Despite the high level of variation, the V4, V5 and V3 sequence variants can be classified by phylogenetic analysis into a small number of discrete lineages. By following the turnover of the different lineages with time, the patterns and process of *env* gene evolution can be examined.

In this chapter, I begin by reviewing these results, and show how the results of these studies can be used to classify the PBMC-derived *env* gene clones to ensure that each of the most predominant V4, V5 and V3 types is represented. In addition, the knowledge of the timing of the expression of the variants in the plasma has been used to determine when sequences very similar to the clones were actively expressed.

Subsequently, a description of the V1 and V2 hypervariable regions of the clones is presented, and the full-sequence of three clones, which have been examined for the presence of any inactivating mutations.

Finally, I will discuss the *in vitro* phenotype inferred from the V3 loop amino acid sequence of each clone.

5.2 A Review of the Sequence Evolution within Patient 82

5.2.1 Analysis of the V4 and V5 Sequence Evolution: the Identification of Distinct Sequence Types

Simmonds *et al*, 1991, obtained sequences of the V4 and V5 regions from serial samples, from both PBMC, and plasma-associated populations, taken over 6 years. The sequences showed a high variation within each sample, including both a variation in amino acid sequence and a variation in length of the hypervariable region. The relationships between the V4 variants, and between the V5 variants was determined by phylogenetic analysis. The V4 sequences grouped into three distinct lineages, named A, B and C. Sequences of the V5 region also clustered into a number of discrete lineages, this time five groups could be distinguished, named A to E, although four sequences fell between the groups and could not be grouped according to this classification.

The grouping based upon the nucleotide sequences was also reflected in the deduced amino acid sequences of the variants. The V4 and V5 regions of each lineage showed a characteristic pattern of amino acid sequence, and of the pattern of sites of potential N-linked glycosylation and also, a discrete pattern of length variation. This is illustrated in table 5.1, which shows the consensus sequence of each of the V4 and V5 lineages, and the sequence of the four unclassified V5 sequences. (Specific examples of the V4 and V5 types can be found in table 5.3, which shows the V4 and V5 sequences of the clones obtained in this study). Within both the V4 and V5 regions, sequences within each lineage rarely differ from each other by more than two or three amino acids. The sites and spacing of N-linked glycosylation sites within the each lineage are relatively constant, as is the variation in length of members within the same group. The length of the V4 consensus sequence, between the conserved flanking regions (of L#FSTW---<V4>---TLPCR), is 16, 17 and 18 amino acids, for V4A, V4B and V4C, respectively. The length of the V5 consensus sequences, between the conserved flanking regions (of DG---<V5>---PGGG) is 10, 11, 15, 11 and 11 amino acids, for the V5 lineages A to E, respectively. As will be

Table 5.1. The Consensus Sequences of the Deduced Amino Acid Sequences of the Different V4 and V5 Lineages

This table is adapted from Simmonds *et al*, 1991.

Residues that are invariant within each lineage are shown in uppercase. Variant residues are shown with the consensus amino acid in lower case. (#) represent sites of potential N-linked glycosylation. Dots have been introduced to allow sequence alignment between the lineages.

opposite page 152

lineage	consensus sequence
V4A V4B V4C	L#FSTW...#StQl#STW#steE#iTLPCR LF#STW.#Ys#gtW?StQhnTeE#ITLPCR LF#STW#STWDLtql#stqnkee#ITLPCR
V5A V5B V5C V5D V5E	DGG.....#gSeTEIFRPGGG DGg....#nTertEiFRPGGG DGGNsG#ks#dTtEtFrPGGG DGGN....rn#TTEiFrPGGG DGG.....DTSNTTEifrPGG
unclassified V5	DGG#...KSKNDPETPRPGGG DGG#.....TSTTEIFRPGGG DGGN....R##TTETRRPGGG DGG#....TSKTTEIFRPGGG

described below, the characteristic amino acid pattern for each of the V4 and V5 lineages was used to classify the cloned variants obtained in this study.

Frequencies and Turnover of the V4 and V5 Types in the Plasma and PBMC Populations

By classifying the V4 and V5 sequence variants into the lineages described above and calculating their frequencies in each year, Simmonds *et al*, 1991, have shown how the viral population changes over time. The frequencies of the V4 and V5 lineages in the plasma, and PBMC populations, at each time point is illustrated in figures 5.1 and 5.2.

V4 Sequence Turnover (figure 5.1)

In the seroconversion sample of 1984, all 9 plasma-derived variants were of type V4A, while in the following sample of 1987, only type V4B were found. In the following sample of 1988, sequence type V4C were most predominant, and had completely replaced type V4B in the two 1989 samples. In 1990 and 1991, type V4C again predominated, but variants of type V4B re-emerged but at low frequency. A similar turnover of sequences was found in the PBMC populations. V4A and V4B sequences were found in 1987, while in the following year, the sequences were almost all of type V4B. Replacement of type V4B sequences with V4C occurred over the subsequent 2 years.

V5 Sequence Turnover (figure 5.2)

A comparable turnover of sequences was found for V5 variants. In the plasma, only V5A sequences were found in the seroconversion sample of 1984. These were replaced by types V5B and V5C in 1987, and by type V5C in the following year. Type V5D and V5E arose in the two 1989 samples, with the continued presence of type V5C sequences. A similar succession of sequence types was found in the PBMC population, although type V5C dominated the data set. V5A, V5B and V5C were found in 1987, with type V5B and V5C predominating in 1988. V5D and V5E, appeared in 1989, to be replaced completely by type V5C in 1990. V5A sequences re-emerged in 1989A although at low frequency.

Figures 5.1(a) and 5.1(b) Frequency of Detection of V4 Sequence Types from Patient 82

a) plasma population

b) PBMC population

Similar histograms were originally published by Simmonds *et al*, 1991. Additional plasma-derived V4 sequences were added to these from the data of Zhang, 1993, and V4 PBMC sequences were added to by the variants sequenced in this study.

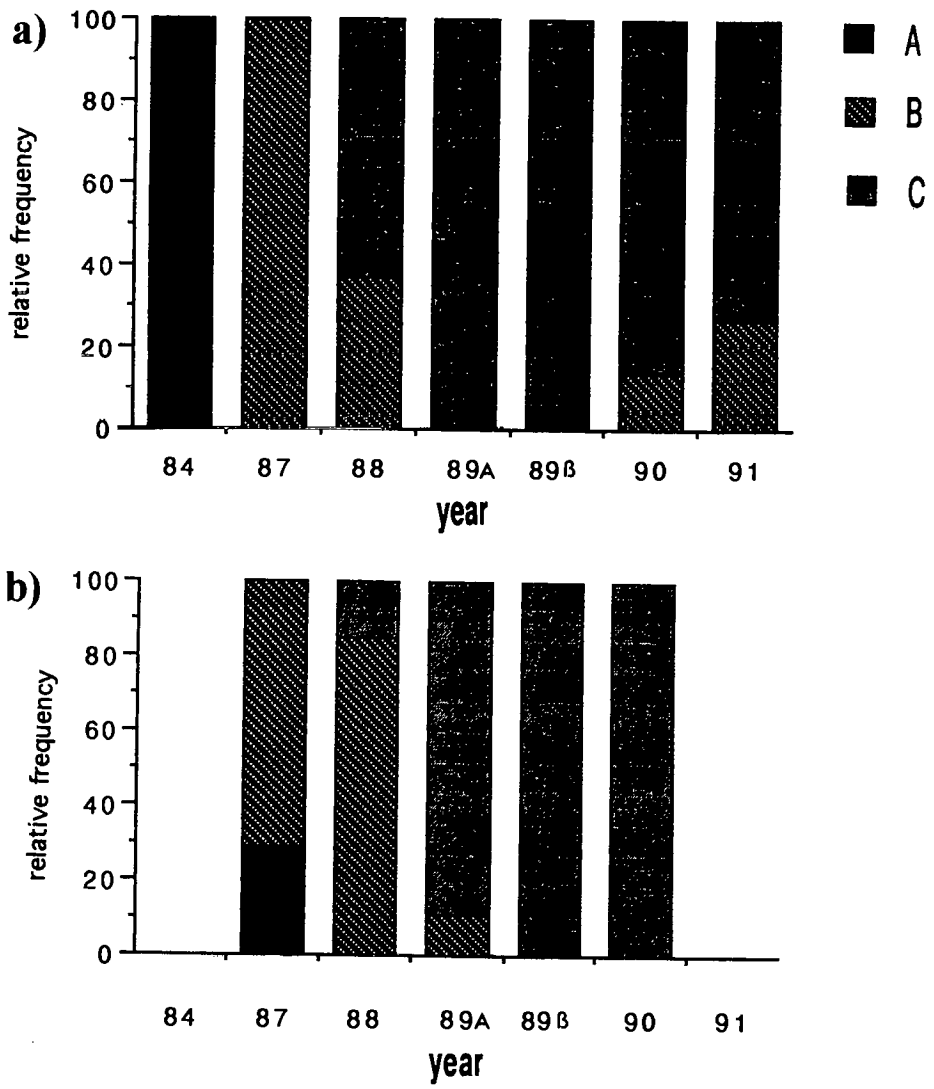
The sample dates are as follows: 1984 (March 1984); 1987 (June 1987); 1988 (January 1988); 1989A (February 1989); 1989B (October 1989); 1990 (April 1990); 1991 (January 1991).

The relative frequencies of each type (A, B and C) was calculated by dividing the number of each sequence type, by the total number of sequences at each time point. The total number of sequences at each time point was:

V4 plasma: 1984 n=09; 1987 n=18; 1988 n=19; 1989A n=13; 1989B n=40; 1990 n=14; 1991 n=11

V4 PBMC: 1987 n=14; 1988 n=13; 1989A n=09; 1989B n=15; 1990 n=03

opposite page 154


a)
100
80
60
40
20
0
relative frequency
84
87
88
89A
89ß
90
91
year
A
B
C
b)
100
80
60
40
20
0
relative frequency
84
87
88
89A
89ß
90
91
year

Figures 5.2(a) and 5.2(b) Frequency of Detection of V5 Sequence Types from Patient 82

a) plasma population

b) PBMC population

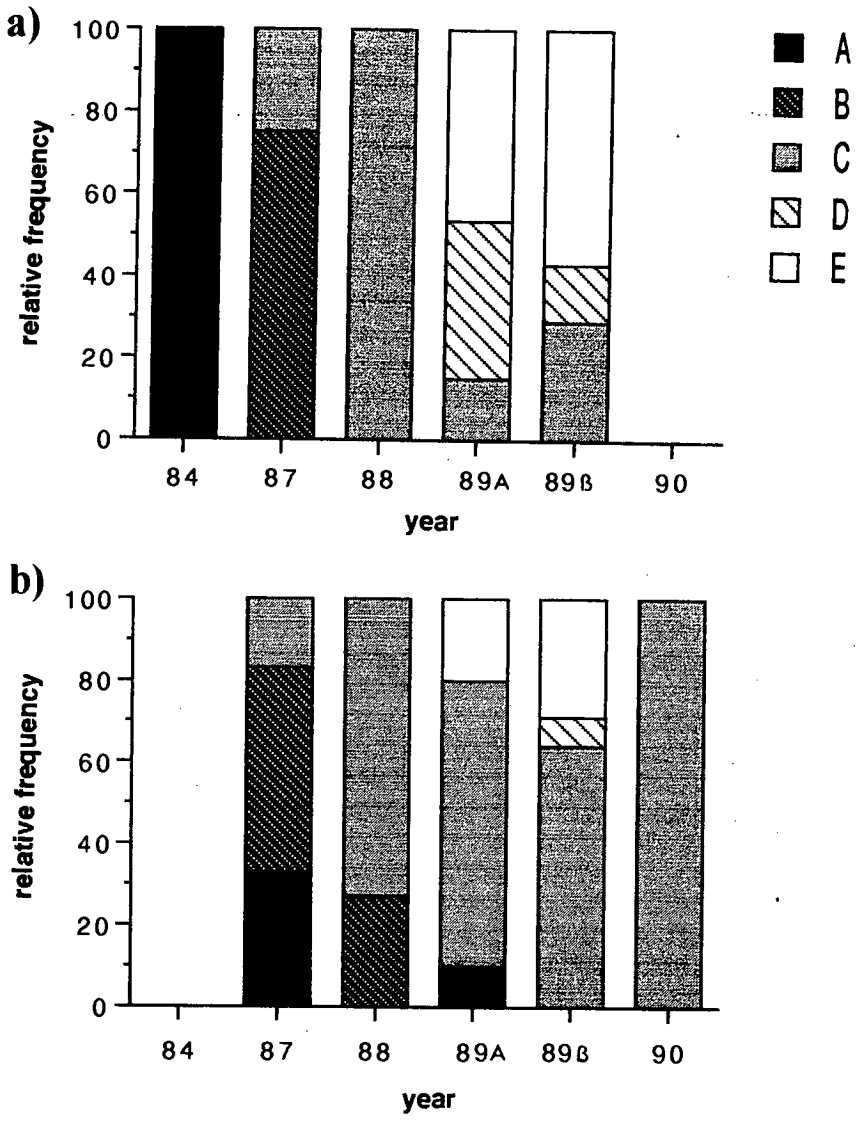
Similar histograms were originally published by Simmonds *et al*, 1991. Additional V5 PBMC sequences were added to by the variants sequenced in this study.

The sample dates are as follows: 1984 (March 1984); 1987 (June 1987); 1988 (January 1988); 1989A (February 1989); 1989B (October 1989); 1990 (April 1990); 1991 (January 1991).

The relative frequencies of each type (A to E) was calculated by dividing the number of each sequence type by the total number of sequences at each time point. The total number of sequences at each time point was:

V5 plasma: 1984 n=02; 1987 n=08; 1988 n=08; 1989A n=13; 1989B n=07
V5 PBMC: 1987 n=12; 1988 n=11; 1989A n=10; 1989B n=14; 1990 n=03

opposite page 155


a)
100
80
60
40
20
0
relative frequency
84
87
88
89A
89ß
90
year
A
B
C
D
E
b)
100
80
60
40
20
0
relative frequency
84
87
88
89A
89ß
90
year

Differences in the Rates of Sequence Turnover in Plasma and PBMC Populations

Simmonds *et al*, 1991, have shown from statistical analysis of the frequencies of the V4 and V5 variants, and from further data derived from the analysis of multiple length variants representing the V4 and V5 types, that there is a significant difference in frequencies of the different V4 and V5 types in the plasma compared to the PBMC population of the same time point. For example, in the 1987 PBMC sample, there were both V4 sequences of type A and B, but in the corresponding 1987 plasma sample, only type B was found (figure 5.1a). Similarly, for V5 region, in the 1987 PMBC sample 3 types V5A, V5B and V5C, could be found, but in the plasma sample only two types, V5B and V5C, were present (figure 5.1b). In 1989A, in the plasma V4E and V5D were of approximately equal frequency, with a low frequency of V5C, however, in the PBMC sample of the same year, type V5C predominated, with a low frequency of type V4E, and a re-emergence of type V5A which was not seen in the plasma in that year.

These differences in the frequencies of variants found in the two populations appears to be the result of two differences in the evolution of the variants in the plasma compared to that in the PBMCs. Firstly, in general, variants appear first, or at a higher frequency in the plasma and only in subsequent years do they reach high frequency in the PBMCs. Secondly, there is a more rapid turnover of sequence variants in the plasma than PBMCs. For example, in the 1987 plasma population V5 types B and C had completely replaced type A found in 1984, however, in the same year, type A, still persisted and formed a substantial proportion of sequences in the PBMC population. Persistence of the V4 seroconversion sequence is also seen in the 1987 PBMC population at a time when it had been completely replaced in the plasma. Similarly, the relative frequencies of types V4B and V4C in 1987, 1988 and 1989 plasma samples compared to the frequencies found in the PBMCs, could be interpreted as a more rapid transition from V4B to V4C in the plasma.

5.2.2 Analysis of V3 Sequence Evolution

Evolutionary Relationships of the V3 Variants

This section reviews the study of Holmes *et al*, 1992, who have analysed the sequence evolution of the V3 hypervariable region in patient 82, over a 7-year period of infection.

A data set of 87 plasma-derived RNA sequences was obtained from six plasma samples, the first taken at the time of seroconversion in 1984, and remaining five from 5 subsequent years of infection (1987, '88, '89, '90 and '91). A phylogenetic analysis was carried out using sequence information from the V3-loop itself, and from 135 nucleotides of flanking sequence, which from its hypervariable nature, provided the majority of the phylogenetic information.

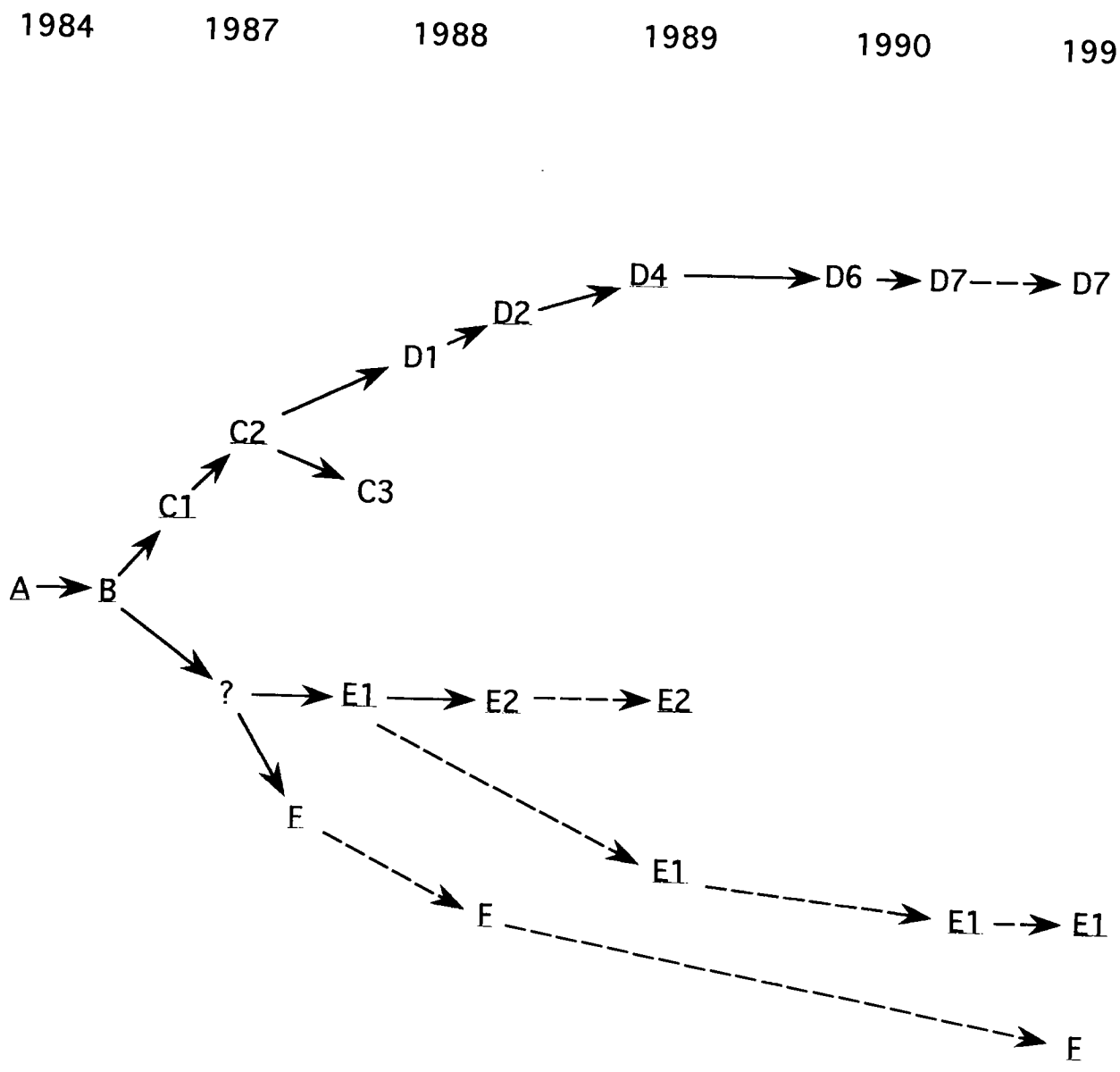
The result of this analysis is depicted in figure 5.3, which shows five principal lineages, denoted A to F. A single sequence (denoted A), was found in all of 12 sequences obtained from the seroconversion (1984) sample. Sequences found three years after seroconversion (1987) had diverged from the seroconversion sequence into a number of discrete lineages (denoted B, C, E and F), which were approximately equidistant from each other. By year four post-seroconversion (1988), two clearly distinct evolutionary lineages (denoted D and E) could be observed, each showing increasing divergence from the earlier sequences. These two lineages predominated for the remainder of the period of infection studied.

Figure 5.3 Evolutionary Framework Relating the Different Amino Acid Sequences found in the V3 Loop from Patient 82

This figure is adapted from figure 2 of Holmes *et al*, 1992.

Distinct evolutionary lineages are designated by letters A to F and sequences within lineages are designated by numbers. Proposed evolutionary relationships are indicated by arrows. Lineages that persist through years are indicated by dashed arrows.

opposite page 158


1984
1987
1988
1989
1990
1991
A
B
C1
C2
C3
D1
D2
D4
D6
D7
D7
?
E1
E2
E2
E
E1
E1
E1
F
F

Classification of the V3 Loop Amino Acid Sequences

Although the phylogenetic analysis used information both within the V3 loop and the variable flanking regions, the sequences within each lineage can be distinguished by their V3 loop amino acid sequence. From the data set of 87 RNA-sequences, only twenty-four amino acid sequences were found. These are listed in table 5.2 according to their lineage; sequences within each lineage are given a number (or 'genotype').

Genotype A, represents the single sequence type found at seroconversion, and is identical to the subtype B consensus sequence, and to the V3 loop sequence of isolates showing macrophage-tropism and a non-syncytium-inducing phenotype.

The divergence of sequences into lineages B, C, E, and F, was characterised by amino acid variation at position 25 of the loop, with glutamic acid (E) of genotype A, replaced by aspartic acid (D) in genotypes B, C1, C4, and C5; glutamine (Q) in genotypes C2 and C3; and glycine (G) in genotype F. Position 25, as will be described in section 5.7 below, is important in determining the *in vitro* viral phenotype. Genotype C4 is of note, in that it contains a non-conservative substitution of arginine (R) to glycine (G), at the tip of the V3 loop, producing 'GPGG', which is rarely found in the Los Alamos data base for subtype B sequences (Myers *et al*, 1991).

Lineages D and E, which predominate in the later stages of infection contain amino acids substitutions that are quite distinct between the two lineages. In particular, the presence of arginine (R) at position 11, valine (V) at 20, (Q) at 25 and asparagine (N) at position 29, distinguish sequences of lineage D. In contrast, sequences of lineage E, each contain glycine (G) at position 11, and the majority contain serine (S) at position 18, and alanine (A) at 23. The serine residue at 18, present in the majority of lineage E sequences produces an unusual 'GPGS' motif at the crown of the V3 loop, which as with the 'GPGG' motif of C3, discussed above, is rarely found in subtype B isolates described in the Los Alamos database (Myers *et al*, 1991). In both of these cases, these sequences reached high frequency in the plasma (Holmes *et al*, 1992), arguing that these sequences can form viable viruses.

Table 5.2 The Amino Acid Sequences of the 24 V3 Loop Genotypes found in the Plasma of Patient 82

This data is taken from Holmes *et al*, 1992. Amino acid sequences are listed according to their evolutionary lineage (A to F). Only residues that differ from those in sequence A (detected at seroconversion) are shown, with a dash denoting identical residues.

opposite page 160

genotype	1 sequence 35
A	CTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHC
B	----------D----------
C1	----------P----------D----------
C2	----------P----------Q----------
C3	----------Q----------
C4	----------P----G------D-T--------
C5	----------D-T--------
D1	----------V----Q----------
D2	----------R--------V---EQ---N------
D3	----------G--------V---EQ---N------
D4	----------R-Y------V---EQ---N------
D5	----------R-Y------V---EQ---N------
D6	----------R-Y------V---DQ----------
D7	----------R-Y------V---DQ---N------
D8	------Y---R-G-----SV--AEQ---N------
E1	----------G------S---A--D----------
E2	----------G------S---A--G----------
E3	----------G------S---A--R----------
E4	----------G------S-V-A--G----------
E5	----------G------S---A--G---N------
E6	----------G------S-V-A--D----------
E7	----------G----------D----------
E8	----------G--------V----D----------
F	----------G----------

Frequency Changes of the V3 Loop Variants

In a similar manner to that described above for the V4 and V5 regions, a number of important observations can be made by examining the changes in frequency of each V3 lineage with time. The frequency changes are given graphically in figure 5.4. These histograms were derived from the data of Holmes *et al*, 1992, for the plasma population, and for the PBMC-population from the data of Simmonds *et al*, 1991, and the clones sequenced in this study.

Firstly, as with the V4 and V5 regions, there is a more rapid turnover in the plasma population than the PBMC population. This results in the continued presence in the PBMCs of variants that had disappeared from the plasma, a phenomenon observed for the V4 and V5 data, described above. For example, genotype A persists until 1987 in the PBMCs, where it makes up 35% (4/11) of the variants sequenced, but genotype A was not found in the plasma at this time point. Similarly, sequences of lineage C persist until 1988 in the PBMC population at a time when lineage C was no longer found in the plasma.

Lineage D and E, which predominated in the later years of infection, alternated in the relative frequency at which they were found in each year. For example, in the 1988 plasma sample, lineage D had the highest frequency, whereas in 1989, lineage E was most predominant. In year 6, 1990, both lineages were of approximately equal frequency, but by year 7, 1991, D again predominated. No new sequence variants were found in year 7, 1991, which coincides with a rapid drop in the CD4 count from this patient, when the immune system may be weakening (Simmonds *et al*, 1991). To explain this, it has been speculated that in the absence of immune selection, fewer new variants will arise.

The frequency fluctuations are paralleled in the PBMC population, but occur one year behind the plasma population, such that lineage E predominates in year 6, 1989, at a time when lineage D predominated in the plasma. However, in 1990, there was an equal distribution of the two lineages D and E as was found in the plasma.

Figure 5.4. Frequency of Detection of V3 Sequence Types from Patient 82.

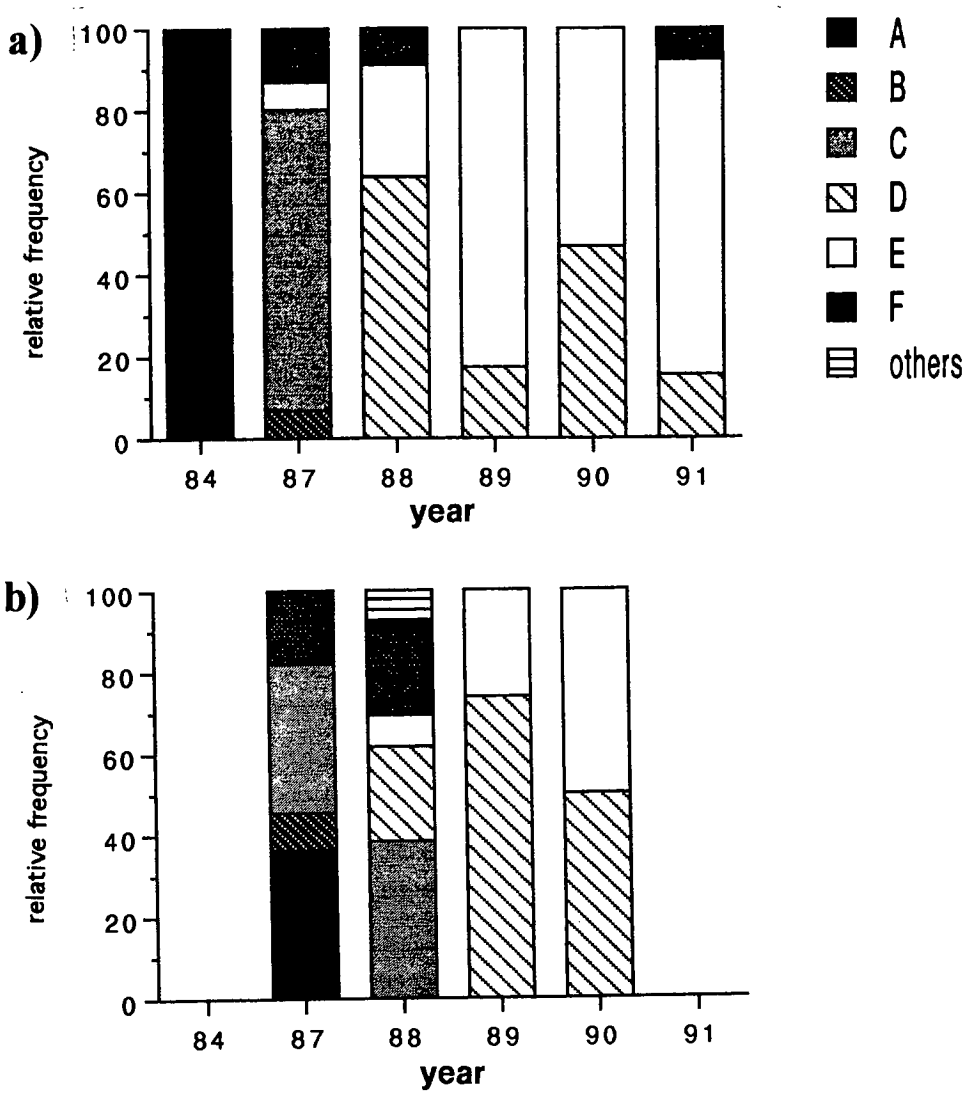
The sequence data used in constructing these histograms comes from Simmonds *et al* 1991, Holmes *et al*, 1992, and the sequences of the clones obtained in this study.

The sample dates are as follows: 1984 (March 1994); 1987 (June 1987)); 1988 (January 1988); 1989 (February 1989 and October 1989); 1990 (April 1990); 1991 (January 1991)

The relative frequency of each sequence type was calculated by dividing the number of each sequence type, by the total number of sequences at each time point. The total number of sequences at each time point was:

plasma: 1984 n=12; 1987 n=15; 1988 n=11; 1989 n=23; 1990 n=15; 1991 n=11
PBMC: 1987 n=11; 1988 n=13; 1989 n=23; 1990 n=10

opposite page 162


a)
100
80
60
40
20
0
relative frequency
84
87
88
89
90
91
year
A
B
C
D
E
F
others
b)
100
80
60
40
20
0
relative frequency
84
87
88
89
90
91
year

The presence over three years of infection, from 1988 to 1991, of two independent and distinct V3 lineages, D and E, is of interest. If these variants show different antigenic properties, then it is possible to speculate that sequences of lineage D, in 1989, were suppressed by the immune system, but remained at low level until they evolved changes at residues that may alter their antigenicity that allowed escape from serum neutralisation, where they could emerge at higher frequency in the subsequent year. Similarly sequences of lineage E, which alternated in frequency with that of lineage D, may have been suppressed in 1988, but evolved amino acid changes that allowed their reaching higher frequency in 1989.

The persistence of sequences such as genotype F, D8 and E1, through a number of years (figure 5.2), may have been possible through the production of amino acid changes outside the loop which altered the antigenicity of the loop. Such mutations mapping outside the V3 loop have been observed to arise in experimental systems McKeating *et al*, 1989; Back *et al*, 1993). Alternatively, they may have been able to persist at low level, where they did not reach a sufficiently high frequency to be countered by neutralising antibodies. Indeed, Holmes *et al*, 1992, showed that the evolution of the V3 region shows some evidence of frequency-dependent selection. Sequences that reached the highest frequency in the plasma showed the greatest reduction in frequency in the next year. For example, genotype A was found at a frequency of 100% 1984, genotype D2 reached a frequency of 45% in 1988, and E2 reached 70% in 1989 . Variants E2 and D2 had disappeared from the plasma by the subsequent year (1989 and 1990, respectively) and genotype A had disappeared by the time of the next sample, 1987. Each of these sequences gave rise to variants with one or two amino acid differences in the V3 loop. These new sequences may confer different antigenic properties on the viruses. However, to test all these predictions and speculations will require the sensitivity to serum neutralisation of the variants to be determined. This is exactly the aim of the work described in this thesis.

Of note, a number (3/13) of unusual V3 loop sequences were found in the 1988 PBMC-population (described as 'others' in figure 5.3). These sequences contained an unusual motif of 'GPGRTV' which was not seen in any of the plasma variants. No attempts were made in this study to express sequences with this motif.

5.2.3 The Implications for this Study

The sequence analysis of Simmonds *et al*, 1991, and Holmes *et al*, 1992, can be used to ensure each of the major *env* sequence types found in the plasma is represented by one or more of the clones obtained in this study. In addition, the V3, V4 and V5 sequence combination of each clone can be used to date when sequences with very similar, or identical, V3, V4 and V5 sequences were present in the plasma. This is particularly important because it has been shown that the presence of a variant in the PBMC population, is not necessarily associated with its active expression in the plasma, and it is the sequences that are present in the plasma that are the targets of the humoral immune response. For later experiments, in which the sensitivity to neutralisation of the variants is determined, the knowledge of the timing of the active expression of each variant will be required to correlate the neutralising sensitivity of each variant, to its first appearance in the plasma. This correlation will be required in order to determine whether the variants arose in response to selection pressure from the neutralising antibody response.

Therefore, although the clones obtained in this study were derived from PBMC DNA, comparison of their V3, V4 and V5 sequences with those sequences derived from the plasma derived variants can be used to show when variants with the same sequence types were actively expressed in the plasma. The classification of the cloned *env* variants, and the determination of when similar sequences were expressed in the plasma will be described in the subsequent sections.

5.3 Classification of the Cloned Variants

5.3.1 Classification of the V4 and V5 Sequences

The V4-C4-V5 amino acid sequences of the *env* gene clones obtained in this study are shown in table 5.3. The clones have been classified into the different V4 and V5 lineages defined above, by comparison of their amino acid sequence to the data set of Simmonds *et al*, 1991 and the more extensive data set of Zhang, 1993.

From the panel of cloned sequences representatives of all 3 V4 types: V4A, V4B and V4C, have been obtained. From the 1987 PBMC sample, A.1 is of type V4A, with the remaining six clones of type V4B. The two clones from the 1988 PBMC DNA sample are V4B and V4C, whilst the clones from the three subsequent PBMC DNA samples (1989A, '89B and '90) are all type V4C.

The clones were more difficult to classify in the V5 region, reflecting the observations of Simmonds *et al*, 1991, in which a number of variants could not be assigned to one of the five V5 lineages (A to F). This in itself reflects the wide diversity of the V5 hypervariable region within p82. However, types V5B, V5C and V5E are represented by the clones. Cloned sequences A.1, C.1, P.2 and E.1 are type V5B; sequence J.1 is type V5E. The remaining clones that could be classified are of type V5C (F.1, U.1, K.1, L.1, M.1, N.1, G.1, I.1 and V.1). Types V5A and V5D were not represented, however, as has been described above, these sequence types did not reach high frequency in the plasma or PBMC population at any time.

Table 5.3. The Deduced Amino Acid Sequences of the V4, C4 and V5 Regions of the PBMC-Derived *env* Clones

The amino acid consensus sequence is given along the top with invariant residues in upper case and variant residues in lower case. The hypervariable regions V4 and V5 are indicated in bold face. The residues that form part of the CD4-binding site (Lasky *et al*, 1987) are indicated by an asterisk (*).

Amino acid residues that differ from the consensus are given. Dashes represent residues identical to the consensus. Dots represent gaps introduced to preserve the alignment and (#) represent sites of potential N-linked glycosylation.

[1] the lineages are assigned according to the classification of Simmonds *et al*, 1991

n.d. not done

opposite page 166

PBMC sample date	clone	<---------V4--------> C4 <----V5----> **F#StW#????????????n?e**ENITLPCRIKQIINMWQeVGKAMYAPPIrGQIRCSSnITGLLLTRDGG**n????????tEt**FRPGGGDM	V4/5 lineage[1]
1987	A.1	-----N.SPQL#NTW..#NT----------T.#....GTEI-------	A/B
	B.1	------YS#DTWSST.QHNT----------#..KSRA#DTEI------	B/U
	D.1	-----NYY#GTW#ST.QHNTG----------..K#ES#TTEI------	B/U
	P.2	------STWDLTQL#STRNK----------T.#.....I--------	B/B
	C.1	-N-Y--YSNG#WSST.QHNT----------G----------#..NT.#.GTEI--------	B/B
	E.1	------YSNG#WTST.QHNT----------#..NT.#.GTEI--------	B/B
	F.1	------YS#GTWIST.QHNT----------GG#KS#DTTET----	B/C
1988	T.1	-----IYS#GTW#PT.QHNT----------#...KSETDPET--------	B/U
	U.1	------STWDLTQL#STQNK----------GG#KS#DTTET--------	C/C
1989A	K.1	------STWDLTQL#STQNK----------K----------SG#KS#DTTEI--------	C/C
	L.1	------STWDLTQL#STQNK----------SG#KS#DTTET---	C/C
	M.1	------STWDLTQL#STQNK----------SG#KS#DTTET---	C/C
	N.1	------STWDLTQL#STQNK----------SG#KS#DTTET--------	C/C
	G.1	------STWDLTQL#STQNK----------S----------SG#KS#DTTET--------	
	R.2	n.d.	
1989B	H.1	------STWDLTQL#STQDK----------K----------SG#KS#DTTEI--------	C/C
1990	J.1	------STWDLTQF...QNE----------DTS#....TTET--------	C/E
	I.1	------STWDLTQL#STQNK----------GG#KS#DTTET--------	C/C
	V.1	------STWDLTQL#STQNK----------K-------T----------SGNKSNDTTEI--------	C/C

5.3.2 Classification of V3 Sequences

The V3 loop amino acid sequences of the clones obtained in this study are given in table 5.4, along with the lineages described by Holmes *et al*, 1992, assigned to them by comparison of their amino acid sequence.

The panel of clones includes a representative of each of the six V3 lineages A to F. Lineages A, B and F, which each contain a single genotype, are represented by cloned sequences A.1, B.1 and F.1, respectively. From lineages C, D and E, which each contain five or more genotypes, the most predominant genotypes are represented: 'C1' is represented by clone T.1; 'C2' by clones D.1 and P.2; 'C3' by clones C.1 and E.1; 'D2' by clones L.1, N.1 and H.1; 'D4' by clones K.1, S.2 and G.1; 'E1' by clone J.1; and 'E2' by clones I.1 and V.1. Cloned sequence M.1 has a D-like V3 loop with a substitution of Q to R at position 25 of the loop - which may be a *Taq*-induced error (chapter 4, section 4.2.2)

Table 5.4. The Deduced V3 Loop Amino Acid Sequence of the PBMC-Derived *env* Clones

The amino acid sequence of clone A.1 is given. Only residues that differ from those in sequence A.1 are shown, with a dash representing identical residues. The genotype (A to F) of each sequence is given according to the classification of Holmes *et al*, 1992.

opposite page 168

PBMC sample date	clone	1 V3 loop 35	genotype
1987	A.1	CTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHC	A
	B.1	----------D----------	B
	D.1	----------P----------Q----------	C2
	P.2	----------P----------Q----------	C2
	C.1	----------G------Q----------	C3
	E.1	----------G------Q----------	C3
	F.1	----------G----------	F
1988	T.1	----------P----------D----------	C1
	U.1	----------G------S---A--D----------	E1
1989A	L.1	----------R--------V---EQ---N------	D2
	M.1	----------R--------V---ER---N------	D-like
	K.1	----------R--------V---EQ---N------	D2
	N.1	----------R-Y------V---EQ---N------	D4
	R.2	----------R-Y------V---EQ---N------	D4
	G.1	----------R-Y------V---EQ---N------	D4
1989B	H.1	----------R--------V---EQ---N------	D2
1990	J.1	----------G------S---A--D----------	E1
	I.1	----------G------S---A--G----------	E2
	V.1	----------G------S---A--G----------	E2

5.3.3 Timing of the Expression of the Cloned Sequences in the Plasma

I have used the V3, V4 and V5 combination of the cloned *env* variants to determine when variants with very similar sequence types were actively expressed in the plasma. Table 5.5 summarises the lineages of each cloned sequence, and dates the expression of similar sequences in the plasma.

Seroconversion (1984) type

Clone A.1 has a V3 type A, which was found only in the plasma seroconversion sample of 1984. The V4 sequence is of type V4A, which again was only found in the 1984 plasma sample. The V5 sequence is of type V5B, which was found only in the 1987 plasma sample. From this analysis, it appears that this clone is very similar to the seroconversion type, but has clearly replicated to produce a 1987 V5B sequence. This V3 loop sequence is also important because it is the subtype B consensus sequence, and is identical, or very similar to, many of the subtype B sequences found *in vivo* from a whole spectrum of patients (Zhu *et al*, 1993; Zhang *et al*, 1993).

Early (1987) types

Clone B.1 has a V3 type B and a V4 type B sequence. Each of these sequence types was found only in the 1987 plasma sample. This therefore dates the expression to 1987 of sequences with very similar V3 and V4 sequences to B.1.

Clones C.1, D.1, P.2, and E.1 have C-lineage V3 sequences, which were only found in the 1987 plasma sample. These four clones have V4 type B, and V5 types - B, or - unclassified. Sequences with this combinations of V3, V4 and V5 types were only found in the 1987 plasma sample.

Clone T.1 has a V3 type C1, found only in the 1987 plasma sample, and similarly, a V4 type B that was found only in the 1987 plasma sample. Therefore, although this clone was obtained from the 1988 PBMC sample, variants with V3 and V4 sequences identical to this were only found at high frequency in 1987.

Clone F.1 has a V3 type F, which was present at low frequency (<15%) in the plasma in 1987, '88 and reappeared, again at low frequency (<10%), in 1990 and 1991. This clone has a V4 type B, and a V5 type C; these types showed the same pattern of expression as the V3 type F, that is, they were present in the plasma in 1987 and '88, and reappeared in 1990 and '91.

Later (1988-91) types:

V3 Loop E1. Clone U.1 has an E1 V3 genotype which was found in the plasma at three time points: 1987, 1990 and 1991. This clone has a V4 type C and V5 type C, which were found in the plasma in 1988, 1989 and 1990. Clone J.1 has the same E1 V3 type as U.1, and a similar V4 of type C. The V5 of J.1 is type E which was found in the plasma in 1989 and 1990. These two clones represent an early and a later version of sequences carrying the V3 type E1. This is in agreement with the dates of the PBMC samples from which they were cloned - U.1 was cloned from the 1988 sample and J.1 from the sample two years later of 1990.

V3 Loop E2. Clones I.1 and V.1 both have E2 V3 genotypes, a V4 type C and a V5 type C. The E2 V3 arose in 1988, but was most predominant in 1989. V4C arose in 1988 and was also found in 1989 and 1990; V5C was found in 1987 onwards. From the V3 sequence, variants very similar to clones I.1 and V.1 were most predominant in the 1989 plasma sample.

D lineage V3 Loops. Clones from the 1989A and 1989B PBMC sample (L.1, M.1, K.1, N.1, R.2 and H.1), each have V4 and V5 type C sequences and a V3 D-lineage sequence. Clones L.1, M.1, N.1 and H.1 have a V3 loop sequence of genotype D2, which was only found in the plasma in 1988. Whereas clones K.1, R.2 and G.1, have V3 genotype D4, which were only found in the plasma in 1989.

With the exception of T.1, U.1 and V.1, each of these cloned sequences has been incorporated into the cassette vector pHXB2-MCSΔ*env* (as described in chapter 4). These 16 cloned sequences are each unique, and represent the breadth of sequence diversity in this patient, analysed by direct sequencing of virus present in the plasma.

Table 5.5 Summary of the V3, V4 and V5 Sequence Lineages of Clones and the Estimated Time of Expression in the Plasma

n.d not done

opposite page 171

5.3.4 Characterisation of the V1 and V2 Sequences

The V1-V2 hypervariable region of 14 *env* gene clones are shown in table 5.6. The V1 region of the clones show both variation in amino acid sequence and in length, similar to the variation shown by the V4 and V5 regions described in section 5.2.1, above. In total, there are eleven different V1 amino acid sequences out of a total of fourteen clones sequenced. Within this region, there are a high concentration of sites of potential N-linked glycosylation, and variation in the number and position of these sites. These sites may influence the conformation of the envelope protein, or mask potential neutralisation sites, and so influence the sensitivity to neutralisation.

No attempts were made to produce a phylogenetic tree from this region due to the extreme variability in length observed between different sequences, which makes sequence alignments unreliable (E. Holmes, personal communication).

The V2 region of the 14 *env* gene clones shows very little sequence variation at either the nucleotide level, or the amino acid level. This is surprising, as variation in the length and number of N-linked glycosylation sites in the V2 region may influence the syncytium-inducing ability of the virus (Groenink *et al*, 1993), and substitutions within the V2 alter the sensitivity to neutralisation by MAbs directed at a neutralising epitope within the V2 loop (Fung *et al*, 1992).

As such, variation in response to selection may be expected in this region, however, we find what appears to be a strong conservation of sequence. The V2 loop (from cysteine to cysteine) is conserved in both length (42 amino acids), and in the number of sites of potential N-linked glycosylation (two, at positions 4 and 34 of the loop). Ten clones show the same amino acid sequence, and four show a single difference in amino acid sequence. P.2 has a glutamine (Q) instead of glutamate (E) at position 16; clones K.1 and N.1 each have a valine (V) at position 5, a site at which all other clones have isoleucine (I); and I.1 has a non-conservative substitution of arginine (R) to methionine (M) at position 10. The unusual conservation of sequence of the V2 region seen with the fourteen clones may be a result of such a small sample-size.

Table 5.6. Deduced Amino Acid Sequences of the V1 and V2 Regions of the PBMC-Derived *env* Gene Clones

The amino acid consensus sequence is given along the top. Invariant residues of the consensus are shown in upper case, with variant residues in lower case.

Residues that differ from the consensus are given in the table, with dashes representing residue identical to the consensus. Dots (.) represent gaps introduced to preserve the alignment. (#) represents sites of potential N-linked glycosylation. The hypervariable regions V1 and V2 are indicated in bold-face in the consensus sequence.
n.d. not done

opposite page 173

PBMC sample date	clone	<----- V1 Region----->	<----------V2 Region---------->
			1 42
		CVTLNCTD**??#aTN?nn????..??Ssge?**MRGEIKN**CSFNiTTSIrDKVQKEYALFYKLDVVPIDEDNTNTSYRLISC**	
1987	A.1	--------WG#---T#S.....----T----------#	#--------
	B.1	--------WG#---I##.....----P----------#	#--------
	D.1	--------LR#---TN#ATNT..#S--W-K----------#	#--------
	P.2	--------LR#---TN#ATNT..#S----E----------#	----------Q----------#--------
	C.1	--------WG#---T#S.....----T----------#	#--------
	E.1	--------WG#---I##.....----P----------#	#--------
	F.1	--------LR#---TN#ATNT..#S--W-P----------#	#--------
1988	T.1	--------LG#---I##.....----P----------#	#--------
	U.1	--------WE#--#ATNT#ATNT#S----K----------#	#--------
1989A	K.1	--------VR#---A##AT#T.T.S.I--E----------#V	#--------
	L.1	n.d	
	M.1	n.d	
	R.2	n.d	
	N.1	--------LR#---A##AT#T.T.S-I--E----------#V	#--------
	G.1	--------VR#---A##AT#T.T.S-I-KE----------#	#--------
1989B	H.1	--------VR#---A##AT#T.T.S-I#-E----------#	#--------
1990	J.1	n.d	
	I.1	--------LR#D--T##AT#I.T.S-I--E----------#	-----M----------#--------
	V.1	n.d	

5.4 Examination of Sequences for Inactivating Mutations

Partially-Sequenced Clones

In total, 20 *env* clones have been obtained from patient 82 in this study. Partial sequence data has been obtained from 18 of these. Of these 18, the majority have been sequenced across each of the five hypervariable regions, and in addition, across the fourth conserved region and across a 200bp section encoding for part of gp41 (tables 4.3, 4.5, 5.3, 5.4, 5.6, and appendix A). In total, 12.47kb of sequence has been obtained from the 18 *env* clones. This is approximately one third of the total length of DNA cloned - 37.2kb (16x1.7kb + 4x2.5kb).

No stop codons or frame-shift mutations have been observed over the 12.47kb of env sequence. This is very low frequency of inactivating mutations and in agreement with Balfe *et al*, 1990, who found one inactivating substitution in 20kb from PBMC-derived *gag* and *env* sequences. Others, using *Taq* polymerase have found a higher frequency.

Although no stop-codons or frameshift mutations could be observed in the panel of clones, one variant, V.1, may contain an inactivating mutation in the C4 region (table 5.3). This variant contains a threonine (T) residue instead of the highly conserved asparagine (N) found in all other patient 82-derived sequences, and indeed, in all published isolates (Myers *et al*, 1991). The highly conserved asparagine suggests that threonine at this site will be an inactivating substitution. This mutation was observed in all 5 clones derived from the amplification of sequence V.1 (139.06), thus it is likely that this mutation was present in the original proviral sequence rather than a *Pfu*-amplification error. Due to this potentially-inactivating mutation, this clone was not expressed in the cassette-vector.

Two other residues within the C4 are variant in the panel of clones obtained in this study (table 5.3). In clone N.1, there is a substitution of glycine (G) in place of glutamate (E), and in clones K.1 and J.1 there is a lysine (K) in place of arginine (R), and in G.1, serine (S) in place of arginine (R) at the same site. However, these sites

are highly polymorphic with a wide range of different residues amongst published isolates. Each of the variant residues found in the clones is represented amongst published isolates, suggesting that these will not be inactivating mutations.

Three Fully-Sequenced Clones

The coding-capacity of three of the twenty clones obtained in this study was determined by sequencing the entire (1.7kb) region of the *env* gene amplified from the PBMC DNA. It was beyond the scope of this study to sequence the entire 1.7kb (or 2.5kb) of amplified sequence from each clone, therefore, I chose to sequence the entire 1.7kb from three clones, A.1, D.1 and G.1. Two of these clones, A.1 and B.1, represent early variants, and the third, G.1, represents a later sequence. In addition, the sequence encoding for almost the all of gp120 and the N-terminal part of gp41 (1.54kb) was obtained from clone P.2, another early variant. This clone, however, contains a 2.5kb fragment amplified from PBMC DNA (see chapter 4, section 4.2.4). Therefore, approximately two thirds of the sequence has been obtained from this clone.

The deduced amino acid sequences of these four clones and for comparison that of the HXB2 sequence, are given in figure 5.5. The three completely sequenced clones (A.1, D.1 and G.1) each contain a full open reading frame, with no frameshift mutations or stop codons. The partial sequence from P.2 also encodes an open reading frame over the region sequenced.

The amino acid sequences of the four clones have been examined for the conserved features common to the envelope proteins of all published subtype B HIV-1 isolates (Myers *et al*, 1991 & '92). The 17 cysteine residues, invariant between highly divergent HIV-1 isolates, are all present in each the clones (in bold-face figure 5.4). In addition, over this stretch of the envelope protein there are 9 highly conserved sites of potential N-linked glycosylation. These sites are conserved in the four clones.

Within the five conserved regions (C1 to C5) of gp120, the clones show amino acid residues characteristic of those of subtype B isolates (Myers *et al*, 1991). Where there are variable residues between the clones, each of the variant residues observed has

been reported in one or more published isolates. For example, at residue 186 there is a valine (V) residue in A.1 and D.1, and an isoleucine (I) residue in G.1 and P.2. Both valine and isoleucine, which are of similar polarity, are found in published isolates, although valine is the most common. This is similarly the case for non-conservative substitutions, for example, at position 207, clone G.1 and HXB2 have a asparagine (N), whereas clones A.1 and P.2, have arginine (R) and D.1 has lysine (K) which are positively charged. Asparagine, lysine and arginine are all seen at this position in published isolates, although asparagine predominates (Myers *et al*, 1991).

There are two exceptions these observations: clones A.1 and D.1, each contain a single site at which there is an amino acid residue, or amino acid insertion, which has never been observed in published isolates. The first is asparagine (N) at residue 253 in clone A.1. This site does, however, show extreme variation between the published isolates (Myers *et al*, 1991). The second exception, is an insertion of two amino acids at 332 to 335, in clone D.1 Variation is seen at this location, however, this particular combination of amino acids has not been reported. It appears to be the result of the duplication of the nucleotide motif TGGAAA.

In conclusion, at this stage of the study, sequence analysis has shown that there are no obvious inactivating mutations, out-of-frame insertions or deletion, nor any in-frame stop codons.

Figure 5.5 Deduced Amino Acid Sequence of the *env* Gene

The amino acid sequence of G.1 is given, with differences from this sequence shown. Dashes represent residues that are identical. Dots are gaps introduced to preserve the alignment.

The five hypervariable regions (V1 to V5) are indicated in boldface. Conserved regions (C1 to C5; according to Modrow *et al*, 1987) are indicated.
represent conserved sites of potential N-linked glycosylation.
Conserved cysteine residues are in boldface: **C**.
The fusion domain of gp41 is indicated.
n.d. not done.

The first amino acid (1) given is at the start of the cassette (5 amino acids down stream of the gp120 signal sequence. The last amino acid (606) is at the end of cassette 1, one third of the way into gp41.

opposite page 177

```
<- - - - - - - - - - - - - - - - C1 - - - - - - - - - - - - - - -
    1
  G.1  TEKLWVTVYY GVPVWKEATT TLFCASDAKA YDTEVHNVWA THACVPTDPN
  A.1  ---------- ---------- ---------- ---------- ----------
  D.1  ---------- ---------- ---------- ---------- ----------
  P.2  ---------- ---------- ---------- ---------- ---------[
 HXB2  ---------- ---------- ---------- ---------- ----------

   51 - - - - - - - - - - - - - - - - - - - - - - - - - - - - ><-
  G.1 PQEVVLE#VT ENFNMWKNNM VEQMHEDVIS LWDQSLKPCV KLTPLCVTLN
  A.1 ---------- ---------- ---------- ---------- ----------
  D.1 ---------- ---------- ---------- ---------- ----------
  P.2 n.d
 HXB2 ------V--- ---D----D- ---------- ---------- -------S-K

  101  - - - - V1 REGION - - - - ->      <- - - - - - - - - -
  G.1 CTDVRNATNA NNATNTTSSI GKEM..RGEI K#CSF#ITTS IRDKVQKEYA
  A.1 ---W..... G-----#--S -ET-..---- -----#---- ----------
  D.1 ---L-----T ------#--S WEK-..---- -----#---- ----------
  P.2    ]-----T ------#--S GEE-..---- -----#---- -------Q--
 HXB2 ---LK..... .-D---#--S GRMIMEK--- -----#-S-- --G-------

  151 V2 REGION - - - - - - - ->><- - - - - - - - - - - C2 -
  G.1 LFYKLDVVPI DEDNT#TSYR LISC#TSVIT QACPKISFEP IPIHYCTPAG
  A.1 ---------- -----#---- --N------- -----V---- ----------
  D.1 ---------- -----#---- ---------- -----V---- ----------
  P.2 ---------- -----#---- ---------- ---------- ----------
 HXB2 F-----II-- -N-..T---S -T-------- ---------- ------A---

  201 - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
  G.1 FAIIKCNDKK FNGTGPCTNV STVQCTHGIR PVVSTQLLL# GSLAEDGVVI
  A.1 --LL--R--- ---------- ---------- ---------- ----------
  D.1 --LL--K--- ---------- ---------- ---------- ------E---
  P.1 --LL--R--- ---------- ---------- ---------- ----------
 HXB2 ---L---N-T ---------- ---------- ---------- -----EE---

  251 - - - - - - - - - - - - - ><- - - - - - - V3 REGION - -
  G.1 RSS#FTDNAK TIIVQL#ESV VITCTRPN#N TRKRIYI..G PGRAVYTTEQ
  A.1 --N------- ------K--- E-#------- ---S-H-..- ----F---GE
  D.1 ---------- ------K--- --#------- ---S-P-..- ----F---GQ
  P.2 ---------- ------Q--- E-#------- ---S-P-..- ----F---GQ
 HXB2 --V------- ------#T-- E-#------- -----R-QR- ----FV-IGK

  301 - - - - -> <- - - - - - - - - - - - - - - C3 - - - - - -
  G.1 IIGNIRQAHC #ISRAKW#ET LKQIVIKLRE QF..E#KTIV FKQSSGGDPE
  A.1 ---D------ #L-------- -R-----L-- --..G----- -N--------
  D.1 ---D------ #--------- -R-----LG- --GNG----- -N--------
  P.2 ---D------ #L-------- -------L-- --..G----- -N--------
 HXB2 .---M----- #-------N- ----DS-L-- --GN.----I ----------

  351 - - - - - - - - ><- - - - - - V4 REGION - - - - -><- -
  G.1 IVMHSFNCGG EFFYC#STQL FNSTW.NSTW DLTQLNST.Q NKEENITLPC
  A.1 --T------- ---------- -----.-... .SP---N-WN -T--------
  D.1 --T------- -----N-A-- -----.-..Y YNGTW---QH NTG-------
  P.2 ---------- -----N---- -----.---- --------.R ----------
 HXB2 --T------- ---------- -----F---- STEGS-N-.. EGSDT-----
```

```
 401 - - - - - - - - C4 - - - - - - - - - - - - - - - - ><- V5 REGION
 G.1 RIKQIINMWQ EVGKAMYAPP ISGQIRCSS# ITGLLLTRDG GNSGNKSNDT
 A.1 ---------- ---------- -R-------- ---------- --.....TNG
 D.1 ---------- ---------- -R-------- ---------- --KNE..SNT
 P.2 ---------- ---------- -R-------- ---------- --.....TNG
HXB2 ---------- K--------- ---------- ---------- ---....-NE

 451 - - -><- - - - - - C5 - - - - - - - - - - - - - ->gp120/gp41
 G.1 TETFRPGGGD MRDNWRSELY RYKVVKIEPL GVAPTKAKRR VVQREKRAVG
 A.1 --I------- ---------- K--------- ---------- ----------
 D.1 --I------- ---------- K--------- ---------- ----------
 P.2 --I------- ---------- K--------- ---------- ----------
HXB2 S-I--L---- ---------- K--------- ---------- ----------

      <-----FUSION DOMAIN------>
 501
 G.2 AIGALFLGFL GAAVSTMGAA AMALTVQARS LLSGIVQQQN NLLRAIEAQQ
 A.1 T---M----- ---G------ S--------L ---------- ----------
 D.1 T---M----- ---G------ S--------L ---------- ----------
 P.2 T---M----- ---G------ S--------L ---------- ----------
HXB2 .-------- ---G------ S--------Q ---------- ----------

 551
 G.1 HLLQLTVWGI KQLQARVLAV ERYLKDQQLL GIWGCSGKLI CTTAVPW#AS
 A.1 ---------- ---------- ---------- ---------- ---T------
 D.1 ---------- ---------- ---------- ---------- ----------
 P.2 ---------- ---------- ---------- ---------- ----------
HXB2 ---------- ------I--- ---------- ---------- ----------

 601
 G.1 WSNKSL
 A.1 ------
 D.1 ------
 P.2 ------
HXB2 ------
```

5.8. Predicted Phenotype of the Cloned Variants

The examination of panels of HIV-1 isolates and genetic mapping studies using recombinant molecular clones of HIV-1, have shown that the V3 loop is a major determinant of viral phenotype. The cellular host range, syncytium-inducing ability, and replicative capacity of HIV-1 strains, have been shown to be determined by the sequence of the V3 loop (O'Brien *et al*, 1990; Shioda *et al*, 1991, Hwang *et al*, 1991).

Fouchier *et al*, 1992, have shown by analysis of a large panel of isolates, that a strong correlation exists between the presence of positively charged amino acid residues at positions 11 or at 25, and a fast-replicating, syncytium-inducing phenotype. It has also been observed that the V3 loop of non-syncytium-inducing (NSI) and macrophage-tropic variants differ little, if at all, from the subtype B consensus sequence. In contrast, syncytium-inducing (SI) and T-cell line-tropic variants that are highly divergent (Chesebro *et al*, 1992).

These two measures of net charge of the V3 loop, and amino acid divergence from the subtype B consensus, can be combined, as described by Milich *et al*, 1993 and Donaldson *et al*, 1994b, to provide an almost complete separation of published isolates into the two groups: SI/T-cell line tropic isolates have a net positive charge of between 4 and 8, and a difference from the consensus of between 5 and 12 amino acids. In contrast, NSI/macrophage tropic isolates show a net positive charge of between 3 and 4, and a difference from the consensus of between 0 and 5 amino acids.

The charge of the amino acid residues at position 11 and 25, described by Fouchier *et al*, 1992, and the combination of net charge of the V3 loop and the number of amino acid differences from the consensus, has been calculated for each clone, in order to predict the phenotype of viruses derived from these clones. These data are given in table 5.7.

All seven clones obtained from the 1987 PBMC DNA sample are predicted to have an NSI/macrophage tropic phenotype, as are the two clones from 1988A, and the three clones from 1990 PBMC DNA samples (table 5.4). Each of these clones contains a neutral of negatively charged amino acid residue at position 11 and 25, and have a similar net V3 charge and number of amino acid differences from the consensus as that associated with NSI/macrophage-tropic isolates. In contrast, all six clones from 1988B and the single clone obtained from 1989, are predicted to be SI and T cell-line tropic. Each of these clones contains the positively charged residue arginine (R) at position 11, and have a similar net charge and number of amino acid differences from the consensus as SI/T cell-line tropic variants.

Of interest, the division of the V3 loop variants according to their predicted phenotype, is also paralleled by the division of variants into the different V3-phylogenetic lineages described above. The sequences that fall into lineages A, B, C, E and F, which are found predominantly in early infection, are all predicted to be NSI/macrophage tropic. The majority of sequences of lineage D, found later in infection (from year 4 post-infection), are predicted to be SI/T-cell line tropic. In addition to the criteria described above that predict an SI phenotype for sequences of lineage D, sequences of this lineage are also characterised by the presence of an asparagine (N) residue at position 29 of the V3 loop (tables 5.2 and 5.4). Asparagine (N) at this position was demonstrated experimentally by De Jong *et al* 1992b, to confer an enhanced syncytium-inducing ability on recombinant molecular clones containing a positive charge at position 306 and/or 320.

The presence, early in the course of infection, of sequences predicted to confer an NSI/macrophage-tropic is in accordance with the studies of Schuitemaker *et al*, 1991 and 1992, who demonstrated that early isolates of HIV-1 are predominantly macrophage-tropic and non-syncytium inducing. Similarly, the co-existence later in infection in patient 82, of both SI/T-cell line tropic and NSI/macrophage-tropic variants, is also in accordance with the work of Schuitemaker *et al*, 1992. In their study, SI isolates were found to emerge later in infection, and that, at these time points, macrophage-tropic isolates could still be isolated, but at lower frequency than SI isolates.

Table 5.7 Analysis of Charge and Diversity of V3 Loop Amino Acid Sequences of Clones Derived from PBMC DNA

Net charge of the V3 loop was calculated by assigning a unitary positive charge to lysine (K) and arginine (R) and a negative charge to glutamate (E) and aspartate (D) residues (see table 5.4).

opposite page 181

year	clone	charge at position:		net charge V3 loop	differences from consensus	predicted phenotype
		11	25			
1987	A.1	0	-1	+3	0	NSI
	B.1	0	-1	+3	1	NSI
	D.1	0	0	+4	2	NSI
	P.2	0	0	+4	2	NSI
	C.1	0	0	+4	2	NSI
	E.1	0	0	+4	2	NSI
	F.1	0	0	+4	1	NSI
1988	T.1	0	-1	+3	2	NSI
	U.1	0	-1	+3	4	NSI
1989A	K.1	+1	0	+5	6	SI
	L.1	+1	0	+5	5	SI
	M.1	+1	+1	+6	6	SI
	N.1	+1	0	+5	5	SI
	G.1	+1	0	+5	6	SI
	R.1	+1	0	+5	6	SI
1989B	H.1	+1	0	+5	5	SI
1990	I.1	0	0	+4	4	NSI
	V.1	0	0	+4	4	NSI
	J.1	0	-1	+3	4	NSI

Chapter 6

Analysis of Virus Infectivity

6.1. Introduction

This chapter describes the results of the transfection of the panel of recombinant proviral clones into cell culture, in order to recover recombinant viruses. This work was carried out in collaboration with Dr Robert Walker and Mrs Elizabeth Harvey, who carried out the manipulations involving infectious virus.

Overview of Transfection Method and Virus Culture

The details of the experimental methods for transfection of proviral DNA and the maintenance of transfected cell cultures can be found in chapter 2, section 2.12. In brief, proviral DNA was introduced by lipofectin-mediated transfection directly into immortalised CD4-positive cell lines, or by transfection into COS-1 cells followed by the addition of the appropriate target cells - either PBMCs from seronegative donors, or immortalised CD4-positive cell lines. Transfected cultures were maintained by the addition of fresh uninfected cells as required, and monitored twice weekly for virus production. In early experiments, virus production was monitored by observing for the production of syncytia. Later experiments were additionally monitored for the production of $p24^{gag}$ antigen, by p24 antigen ELISA of culture supernatants, and by immunofluorescent staining of cells with an anti-p24 monoclonal antibody.

Four immortalised CD4-positive T-cell lines were used for transfection and virus propagation: H9, C8166, SupT1 and $Jurkat_{tat}$; and one immortalised CD4-positive cell-line of myeloid origin: U937. PHA-stimulated PBMCs obtained from seronegative donors were also used (chapter 2, section 2.11).

6.2. Results

I will describe in detail the time course of a representative transfection experiment. Subsequently, I will present the data from the transfection experiments involving the constructs generated in this study by summarising these parameters. In addition, where appropriate, the full data from a number of other transfections will be presented in order to illustrate specific points.

6.2.1 Data from a Representative Transfection Experiment

In this example, four PBMC-derived recombinant proviral clones: E.1, F.1, G.1, H.1, and the positive control pHXB2-MCS, were transfected into COS-1 cells, followed by co-cultivation with PBMCs at day 3. Figure 6.1 and table 6.1, show the data for supernatant p24 antigen production and for the presence of p24-positive staining cells cells.

At 72hr post-transfection the transfected COS-1 cells were assayed for the presence of p24-positive staining cells (table 6.1). At this time point, all transfected COS-1 cell cultures contained p24-positive-staining cells, indicating that each transfection had successfully resulted in the uptake and expression of the proviral DNA. The efficiency of the transfection was in the range of 5.7% to 10 %, as determined by the percentage of p24-positive-staining COS-1 cells 72hr post-transfection. At 72hr post-transfection, culture supernatants were also monitored for the level of p24 antigen (table 6.1). An initial high burst of p24-production was observed with each of the transfected cultures and lasted 1 to 2 days, again indicating the uptake and expression of the proviral DNA. The amount of p24 produced from the transfected COS-1 cells varied widely with each transfected culture. For example, a relatively large amount of p24 was produced from the transfections with clones E.1 and F.1, with 23 ng/ml and 56 ng/ml, respectively (table 6.1). The control transfection with pHXB2-MCS, however, produced only 7 ng/ml of p24 antigen. These values do not correlate with the percentage of p24-positive cells detected at the same time point.

Figure 6.1 and Table 6.1 p24-Antigen Production after Transfection into COS-1 Cells followed by Co-Cultivation with PBMCs for clones pHXB2-MCS, E.1, F.1, G.1 and H.1

Figure 6.1 gives the time course of p24-antigen production in culture supernatants
Table 6.1 shows the presence p24-positive staining cells in the transfected cultures.

+ the presence of p24-positive cells at (> 10%)
w+ cultures with less that 10% cells staining for p24 antigen
- cultures with no p24 staining above background

opposite page 185

Figure 6.1

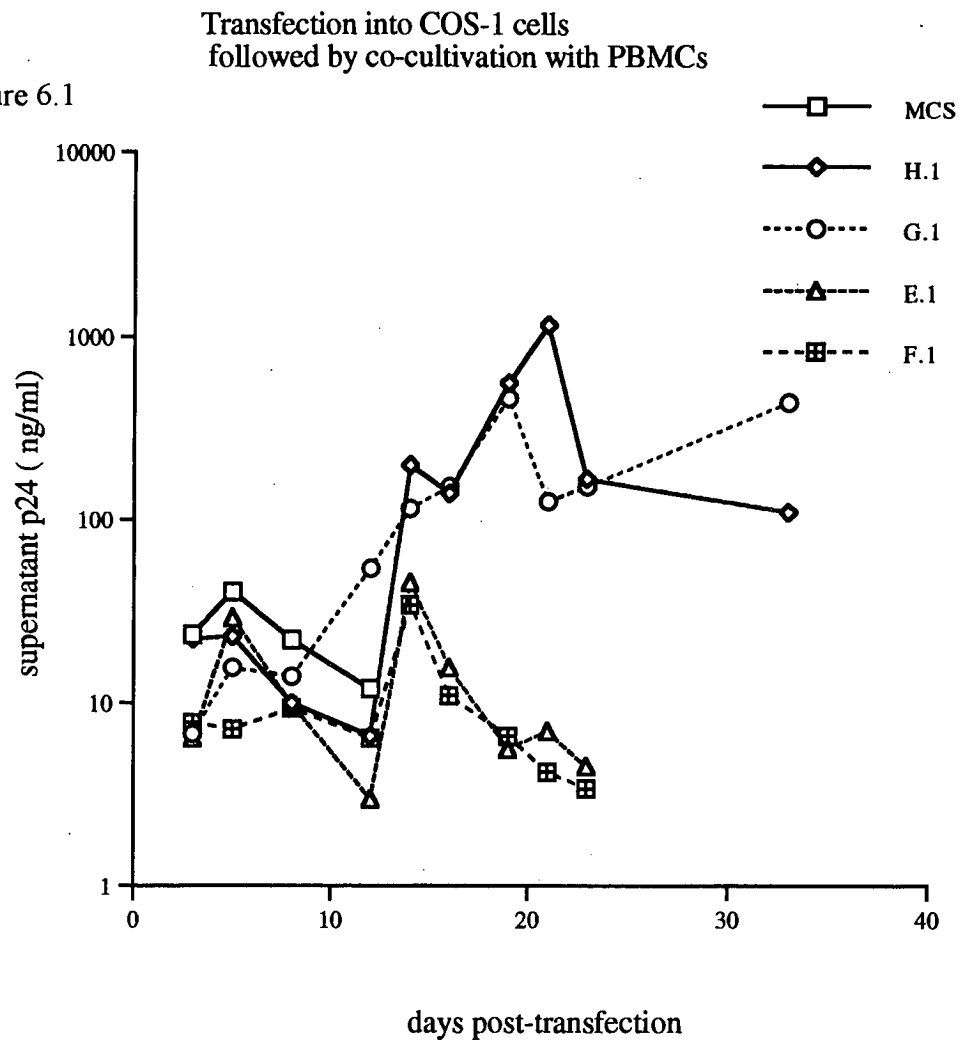


Table 6.1.

			days post transfection									
clone	72hr[1]	72hr[2]	3	5	8	12	14	16	19	21	23	33
MCS	+ (10.0)	7.0	+	+	+	+	culture stopped					
H.1	+ (7.9)	9.0	+	+	+	+	+	+	+	+	+	+
G.1	+ (7.0)	9.4	+	+	+	+	+	+	+	+	+	+
E.1	+ (6.2)	23.0	+	w+	w+	w+	-	-	-	-	-	-
F.1	+ (5.7)	56.0	+	w+	w+	-	-	-	-	-	-	-

[1] COS-1 monolayer, figures in brackets indicate the percentage of COS-1 cells staining positive for p24.

[2] p24 (ng/ml) in supernatant from the COS-1 transfected monolayers.

Following the addition of PBMCs to the cultures, it was not possible to determine accurately the percentage of p24-positive cells on account of the mixed population of cells within the PBMC cultures. Therefore, following the co-cultivation of PBMCs p24-positive cells were scored by indicating either the presence at greater than 10% of p24-positive cells (+), or the absence of p24-positive cells (-). Cultures that showed only a few p24-positive cells (less than 10%) were recorded as weakly positive (w+).

In the example in table 6.1, cultures from the three clones pHXB2-MCS, G.1 and H.1, consistently showed p24-positive cells throughout the period of culture monitored. For these three clones, the results of the p24-cell staining were paralleled by the production of p24 in the culture supernatants. Each culture produced an increasing amount of p24 over the course of culturing, with a maximum production of 1160 ng/ml for H.1, 460 ng/ml for G.1, and 54ng/ml for pHXB2-MCS. Virions from these clones were considered infectious, as there is a clear indication from these data of virus infectivity and replication. Confirmation of the infectious nature of the virions recovered from the clones was carried out by demonstrating cell-to-cell transmission of virus by co-culturing the transfected cultures with uninfected cultures of the T-cell line C8166. Following co-cultivation with these clones increasing levels of p24 antigen could be detected in the culture supernatant, and p24-positive cells could be observed (data not shown).

Virions from the other two PBMC-derived clones, E.1 and F.1, show indications of replication in the PBMCs in the early stages of culturing. p24-positive cells could be detected at day 3 post-transfection, but the number of p24-positive cells slowly declined over time, until only a small number of cells stained p24-positive, at days 5, 8, and 12 for E.1, and day 5 and day 8 for F.1 (table 6.1). p24-positive cells could not be detected on day 14 and thereafter from these two clones. Similarly, p24 detection in the supernatant from these cultures, showed an increase between day 12 and day 14, but subsequently, p24 levels showed a steady decline over time (figure 6.1). The presence of infectious virions from these two clones could not be demonstrated when cells taken on day 6 post-transfection were co-cultivated with C8166 cells; no p24 could be detected in the culture supernatant and no p24-positive

cells could be observed over a two-week period of culturing monitored (data not shown).

6.2.2. Transfection of an *env*-Defective Proviral Construct pHXB2-MCSΔ*env*

In order to facilitate the interpretation of results, three transfections with the env-defective proviral construct pHXB2-MCSΔ*env* were carried out in order to determine the background levels of p24 production resulting from the expression of a defective provirus.

Four transfections were carried with this construct: three directly into H9 or C8166 cultures, and third into COS-1 cells followed by co-cultivation with H9 cells (table 6.5(iv)). The first three transfections were monitored by observing for the production of syncytia. In each case, no syncytia were observed over a three-week period of monitoring the cultures, indicating that this clone does not produce viable virions.

The fourth transfection, which was into COS-1 cells followed by co-cultivation with H9 cells, was monitored by assaying for the production of p24. The data from this experiment are given in figure 6.2 and table 6.2. The initial transfection produced high levels of p24 (24ng/ml) in the supernatant and 2.8% p24-positive staining COS-1 cells. Following the addition of H9 cells, p24 levels in the supernatant declined steadily over time to reach <1ng/ml at day 22. No p24-positive H9 cells could be detected at any of the time points post-transfection. In a simultaneous transfection, the parental clone pHXB2-MCS produced sustained levels of p24 in the culture supernatant and p24-positive staining H9 cells.

The results of this experiment would suggest that transfection of an *env*-defective provirus into the cell-lines used in this study (COS-1, H9 and C8166), can result in the expression of p24 antigen, and production of virions, as indicated by the presence of p24 antigen in the supernatant. However, levels of p24 in the supernatant decline

steadily with time, and no p24-positive staining cells can be detected in cells added to the cultures after transfection. This indicates that any virions produced are not infectious.

By extrapolating this result to other transfections, such as in the example given above for clones E.1 and F.1, the presence of p24-positive PBMCs implies that virions from these clones had indeed infected the target PBMCs, and that they are minimally infectious.

Figure 6.2 and Table 6.2 p24-Antigen Production after Transfection into COS-1 Cells followed by Co-Cultivation with H9 cells for clones pHXB2-MCS and pHXB2-MCSΔ*env*

Figure 6.2 gives the time course of p24-antigen production in culture supernatants Table 6.2 shows the presence p24-positive staining cells in the transfected cultures.

+ the presence of p24-positive cells at (> 10%)

w+ cultures with less that 10% cells staining for p24 antigen

- cultures with no p24 staining above background

opposite page 189

Figure 6.2

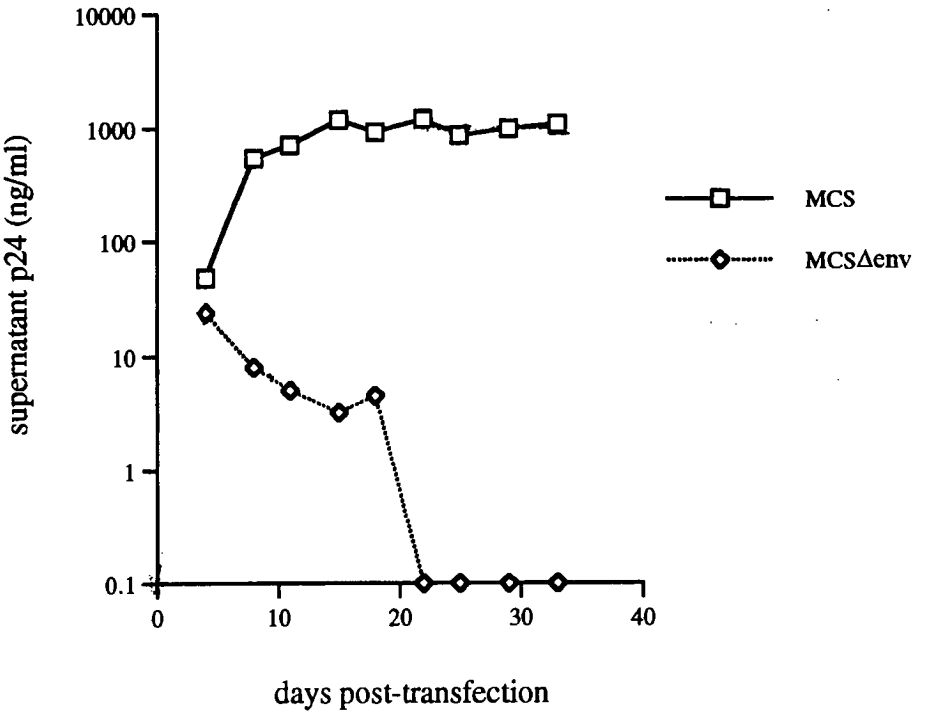


Table 6.2

	days post transfection										
clone	72hr[1]	72hr[2]	4	8	11	15	18	22	25	29	33
MCS	+ (11.4)	48.0	+	+	+	+	+	+	+	+	+
MCS Δenv	+ (2.8)	24.0	-	-	-	-	-	-	-	-	-

[1] COS-1 monolayer, figures in brackets indicate the percentage of COS-1 cells staining positive for p24.

[2] p24 (ng/ml) in supernatant from the COS-1 transfected monolayers.

6.2.3 Transfections of Controls pHXB2-SPT and pHXB2-MCS

To assess the ability of our transfection procedure to recover infectious viruses, we conducted each transfections experiment in parallel with one, or both, of two control proviral constructs pHXB2-SPT and pHXB2-MCS, previously demonstrated to produce infectious virus (chapter 3, section 3.5.).

Over the course of this study, a total of thirty-two transfections were carried out with either pHXB2-SPT or pHXB2-MCS (table 6.5(iv)).

Transfections into T-cell Lines

A total of twenty transfections with were carried out with pHXB2-SPT and pHXB2-MCS, directly into four T-cell lines: H9, C8166, Jurkat$_{tatIII}$, and SupT1, or indirectly into COS-1 cells followed by the addition of H9 or C8166 cells (table 6.5 (iv)).

Sixteen out of the total of twenty transfections with these two clones resulted in the production of viable virus as determined by the presence of syncytia in the culture flasks for early experiments, and by the production of p24 antigen from the later experiments. In each case syncytia were detected within one week of transfection.

The four transfections which failed to result in the production of syncytia were four early transfections. These initial failures were probably due to the failure to introduce the proviral DNA into the target cells, when we were optimising transfection conditions.

The efficiency of transfection, measured as the rate and level of p24 production, varied widely with each transfection. Peak p24 levels reached a maximum of greater than 1000ng/ml with each successful transfection, but due to variation in the rate of

p24 production, these levels were reached between day 10 and day 30 post-transfection.

Transfections Directly into PBMCs

Five transfections of pHXB2-SPT and pHXB2-MCS were carried out directly into PBMC cultures (table 6.5 (iv)). Only one of these transfections was successful: pHXB2-MCS produced infectious virions, as determined by the presence of syncytia at day 22 post-transfection. In consequence, the method of first transfecting into COS-1 cells followed by co-cultivation with PBMCs was used to recover virions in PBMCs for all subsequent experiments.

In seven transfections into COS-1 followed by co-cultivation with PBMCs, virions were successfully recovered from six (table 6.5 (iv)). The single failure may have due to the variability of different batches of PBMCs to support replication of these viruses (Yamada *et al*, 1990; Spira and Ho, 1995; Iverson *et al*, 1995). As a consequence, all batches of donor PBMCs were tested for their ability to support viral replication of the standard laboratory isolates HIV-1_{IIIB} or HIV-1_{MN}.

Variation between Individual Transfection Experiments

Although the majority of transfections into COS-1 cells followed by co-cultivation with PBMCs for clones pHXB2-MCS and pHXB2-SPT were successful in recovering virions, there was considerable variation in the rate of p24 antigen production. In some transfections p24 could be detected from day 3 post-transfection, whilst in others, p24 could not be detected until week three post-transfection. Peak p24 levels following virus propagation were generally in the range of 50 to 75ng/ml, an order of magnitude lower than in T-cell lines (data not shown).

These variations in the rate of p24 production presumably reflect variation in the quality of DNA, the efficiency of the proviral DNA uptake, or the quality of the cell

cultures. In light of the variation of p24 production after transfection, a number of repeat transfections were carried out with many of the constructs generated in this study.

6.2.4. Transfection of PBMC-Derived Recombinant Proviral Clones

The initial transfection experiments were carried out with recombinant proviral clones constructed by replacing a 1.7kb region (cassette-1) of the *env* gene of pHXB2-MCS, with sequences cloned from PBMC DNA. From the results of these experiments, the infectivity of other constructs, generated by the exchange of a 2.5kb fragment (cassette-2), and 1.4kb region (cassette-3) with PBMC-derived sequences, was determined.

Recombinant Clones Constructed by Exchange of Cassette-1

At the start of the transfections conducted in this study, much of the published work characterising the growth properties of HIV-1 isolates concentrated on laboratory strains, such as the isolates HIV-1_{IIIB}, HIV-1_{MN}, and HIV-1_{SF2}. These isolates were favoured because they are capable of growth to high titre in transformed CD4+ cell lines.

In a similar manner, due to the ease of handling transformed cell lines compared to primary cell cultures, the early transfections in this laboratory were carried out into transformed cell lines. In addition, the detection of virus production in transformed cell lines can be easily monitored, by observing for the production of syncytia.

Transfections into T-cell-Lines

The first transfections were carried out with the ten recombinant proviral clones

whose *env* genes were amplified from PBMC DNA using *Pfu* polymerase: A.1 to J.1 (table 6.5 (i)).

A total of thirty-two transfections were carried out with these ten recombinant clones into five different immortalised cell-lines: C8166, H9, SupT1, Jurkat$_{tatIII}$ and U937. All transfection experiments were set up in parallel with one or both of the control constructs pHXB2-SPT and pHXB2-MCS. A given set of transfection experiments was considered successful when virus could be recovered from one or both of the control clones.

Results of Transfection into T-cell Lines

From the early transfections in which virus production was monitored by observing for the presence of syncytia, only two of the ten proviral clones (G.1 and H.1) produced syncytia. The remaining eight clones failed to produce syncytia despite numerous attempts with many of the constructs and the use of several different immortalised cell lines (table 6.5 (i)).

In later experiments, where virus production was monitored by measuring p24 production, each of the ten proviral clones produced an initial burst of p24 (ranging from between 24.0ng/ml and 44.0ng/ml) into the culture supernatant, and p24-positive cells (at greater than 10 %) could be detected 72hr post-transfection. However, sustained p24 production was only observed with clones G.1 and H.1. The other eight clones showed declining levels of p24 in the culture supernatant, which dropped below the level of detection (<1ng/ml) within 18 days, and no p24-positive staining cells could be detected from day 3 post-transfection onwards.

Syncytia were observed in all cultures transfected with G.1 and H.1, but no syncytia were observed in cultures transfected with the other eight clones. The first appearance of syncytia was slightly delayed in cultures transfected with G.1 and H.1, as compared to the HXB2 controls. For example, from a set of transfections into C8166

cells, syncytia were detected at day 11 for H.1, but at day 6 with pHXB2-SPT, and at day 5 with pHXB2-MCS. Similarly, in a second transfection experiment, syncytia were detected at day 22 in the H9 cells transfected with H.1, but at day 7 for pHXB2-MCS.

The infectivity of the virions obtained from G.1 and H.1 was confirmed by passaging aliquots of the transfected cultures to fresh uninfected C8166 cells, and by initiating infections by inoculating C8166 cultures with cell-free supernatants obtained from the transfected cultures. In each case, virus was observed to spread and replicate in these cultures as determined by an increase in p24 production with time, and the presence of p24-staining cells (data not shown).

All attempts to transmit virus from the other eight clones by cell-to-cell passage, or through inoculation with cell-free supernatants into C8166 cells, failed to demonstrate the presence of infectious virus: no p24 antigen could be detected, above the background levels from the inoculum, and no p24-positive cells could be observed (data not shown).

Transfections with PBMCs

At this stage of the study it was becoming apparent from studies taking place in other laboratories, that many primary isolates are not able to grow in transformed CD4+ cell lines (Fenyo *et al*, 1989; Tersmette *et al*, 1989b; Schuitemaker *et al*, 1991). In these studies, the primary isolates which were unable to grow in transformed CD4+ cell lines showed a preference for growth in CD4+ T-lymphocyte cultures, or cultures of monocyte-derived macrophages. Many such isolates replicated slowly and to low level in primary cell culture and failed to infect T-cell lines productively. Those that did grow in T-cell-lines did not induce the production of syncytia in these cell types, nor in T-cell cultures or in macrophages.

At the same time, the differences in the *in vitro* growth properties of HIV-1 isolates were found to be determined by the *env* gene. These different growth properties could

be transferred to recombinant viruses by substitution of the relevant *env* genes, or gene-fragments (Westervelt *et al*, 1991; Cann *et al*, 1992; O'Brien *et al*, 1990).

By combining these findings, we hypothesised that many of the PBMC-derived *env* genes obtained in this study would confer growth properties similar to the slow-replicating, non-T-cell-line tropic isolates described. It was perhaps not surprising, therefore, that infectious virions could not be recovered from the majority (8/10) of PBMC-derived clones when transfecting into T-cell lines.

The study of Fouchier *et al*, 1992, and Milich *et al*, 1993, correlating the presence of positively-charged V3 loop with a fast-replicating, T-cell-line tropic syncytium-inducing viral phenotype, allowed us to make more precise predictions about the growth properties of our recombinant proviral clones. The growth properties inferred from the charge and amino acid sequence of the V3 loop, has been described in the previous chapter (chapter 5, section 5.5.). Eight out of the ten PBMC-derived clones are predicted to confer a slowly replicating, non-syncytium-inducing, non-T-cell-line tropic phenotype, whilst two clones, G.1 and H.1, are predicted to confer a fast-replicating, syncytium-inducing, T-cell-line tropic phenotype (table 6.5(i)).

The results of the transfections into immortalised cell-lines are compatible with these predictions. The production of infectious virions was observed only in cultures transfected with the two predicted-SI clones, G.1 and H.1. No evidence of virus infectivity could not found from cultures transfected with the eight predicted-NSI clones.

In the light of this, it appeared likely that culture in donor PBMCs would be required to recover and replicate many of the PBMC-derived clones constructed in this study. Subsequent transfections were therefore carried out with PHA-stimulated PBMCs as the target cells.

Results of Transfections into PBMCs

Each of the ten proviral clones A.1 to J.1, was transfected into COS-1 cells, followed by co-cultivation with PBMCs, as summarised in table 6.5(i). In each case, an initial burst of p24 antigen in the culture supernatant (ranging from between 1.8ng/ml to 38ng/ml) was produced, indicating that each transfection had been successful (data not shown). However, sustained p24 antigen production over the four-week course of infection monitored was only observed from the two clones, G.1 and H.1, previously demonstrated to replicate in T-cell-lines. For the other eight clones, p24-positive staining PBMCs were not detected at any of the time points monitored, and p24-antigen levels in the culture supernatant dropped to below 1ng/ml within one week of culturing (data not shown). Virus from the two clones, G.1 and H.1, could be transmitted to the cell-line C8166. This was demonstrated by the production of syncytia, the detection of p24 antigen in the supernatant, and the presence of p24-positive cells two days after co-cultivation of C8116 cells with the transfected PBMCs. The co-cultivation of C8166 cells with the PBMCs taken from the transfections of the remaining eight clones, as perhaps expected, did not show any evidence of virus replication; no syncytia were observed in the co-cultivated C8166 cells and no p24-positive cells were detected, nor was any p24 antigen produced into the supernatants (data not shown).

Due to the variability between transfections, we conducted two further transfections of three of the clones A.1, E.1 and F.1, that did not produce evidence of virus replication on the first attempt propagating with PBMCs. All attempts with clone A.1 failed to recover infectious virions; p24 could be detected in the culture supernatant immediately following transfection, and similarly >10% of COS-1 cells stained positive for p24. However, p24 production declined over the course of infection and no p24-positive staining PBMCs were found.

In the two repeat transfections with F.1, and in one of the two repeats with E.1, there was an indication of the production of infectious virions from each clone. One of these transfections has been described in section 6.2.1 above, given as an illustrative example of the time course of a transfection experiment. In this transfection, p24

positive PBMCs could be detected up to day 12 with E.1 and up to day 8 with F.1, implying that early on in the course of the transfection, virions from these clones was capable of infecting PBMCs. Similarly, from examination of the p24 in the supernatant, there was an increase in p24 production between days 12 and 14, which indicates the presence of actively-replicating viruses. The second repeat transfection with these two clones will be described below, in section 6.2.6.

In summary, with the ten *Pfu*-amplified proviral clones constructed through exchange of cassette-1, we were able to recover infectious virions from two clones, G.1 and H.1. Of the remaining eight proviral constructs, six (A.1, B.1, C.1, D.1, I.1 and J.1), clearly failed to produce infectious virions, and two (E.1 and F.1), showed some evidence early on in the course of transfection of low-level infectivity.

Taq-Polymerase Amplified Clones

The two infectious clones, G.1 and H.1, each have a V3 loop sequence predictive of a T-cell line-tropic/SI-ability. The remaining eight proviral constructs each have V3 loops predictive of an non-T-cell-line-tropic/NSI-phenotype. It could be hypothesised that a predicted SI phenotype is required for the production of infectious virions in the pHXB2-MCS expression system that we have developed. If this was so, it would be expected that transfections of further SI variants should produce viable virus.

The only proviral clones available at this stage of the study, predicted to confer an SI phenotype, were those amplified from PBMC DNA with *Taq* polymerase K.1, L.1, M.1, and N.1 (table 6.5(i)). We therefore conducted a series of transfections with these constructs, despite the high error rate demonstrated to result from amplifications with this enzyme, which may have introduced errors into these clones (chapter 4, section 4.2.2).

Results of Transfections with *Taq* Amplified Clones

A single transfection was carried out into COS-1 cells, followed by co-cultivation

with PBMCs, with each of the four *Taq*-derived constructs (table 6.5(i)). However, no evidence of virus replication could be found for any of the clones. There were clear signs of p24 production following the initial introduction of the DNA into the COS-1 cells; both p24-positive-staining cells and p24 antigen in the supernatant could be detected. However, following the addition of the PBMCs, the p24 in the supernatant declined to <1ng/ml over a period of one week, and no p24-positive cells could be detected from the PBMC cultures at any time points. Transfections conducted with the control pHXB2, or G.1 and H.1, showed increasing levels of p24 in the supernatant, and p24-positive cells could be observed, indicating that transfection procedures and virus propagation were successful.

Similarly, transfection of the four *Taq*-derived constructs with T-cell lines H9 or C8166 as the propagating cells, failed to recover infectious virus (table 6.5(i)); no p24 could be detected in the supernatants, and no p24-positive cells were observed beyond those detected immediately following transfection (data not shown).

Possible Inactivating Mutations in *Taq* Amplified Clones

Despite the predicted SI phenotype of these clones, each failed to produce evidence of virus replication following transfection into PBMCs or T-cell lines. These clones were obtained from the same PBMC DNA sample as the infectious clone G.1 and contain very similar *env* gene sequences to both G.1 and H.1 (chapter 5, section 5.3.2). The most obvious explanation for the failure to recover infectious virus is the difference in the amplification procedures used to obtain these clones. The non-infectious clones were amplified with *Taq* polymerase, which was demonstrated to introduce errors into the *env* sequences during amplification. The infectious clones G.1 and H.1 were amplified with *Pfu* polymerase, an amplification procedure from which I could not demonstrate the introduction of errors. It is conceivable, therefore, that clones K.1, L.1, M.1, and N.1 contain inactivating mutations introduced during amplification. However, further analysis would be required to confirm this.

6.2.4 Transfections with Additional PBMC-Derived Constructs Exchanging Cassettes -2 and -3

From the results of the transfections described so far, we have hypothesised that *env* sequences of an NSI-predicted phenotype are unable to form fully-infectious viruses in the pHXB2-MCS expression system. In contrast, we have demonstrated that, with the exception of the *Taq*-amplified sequences, predicted SI-sequences are able to form infectious viruses.

One reason for the failure to produce productive infections from predicted-NSI sequences is each of the eight clones (A.1 to F.1, I.1 and J.1) contains an inactivating mutation. However, we did not detect any inactivating mutations following extensive sequencing of the clones. Two of these clones (A.1 and D.1) have been fully-sequenced and shown to have an intact reading-frame, with no obvious inactivating mutations (chapter 5, section 5.8).

So we decided to pursue further the hypothesis that there may be something specific about inferred-NSI sequences which prevents their production of fully-infectious viruses. Firstly, we made two more constructs with predicted NSI-sequences using the other two cassettes, cassette-2 which contains the entire *env* gene, and cassette-3 which encodes only for gp120. Secondly, we constructed three more SI-clones exchanging cassette-2 to examine whether swapping gp160 had any noticeable effects on growth kinetics. Finally, we aimed to construct recombinant clones using the *env* genes from two clones of known viability and phenotype, $SF2_{MC}$ and $SF162_{MC}$, which are SI and NSI, respectively (Cheng-Mayer *et al*, 1989 & 1990).

Construction of Clones P.2 and A.3

Clone P.2 has the 2.5kb fragment (cassette-2) of the vector has been replaced by an homologous fragment amplified and cloned from PBMC DNA (chapter 4). The V3 loop sequence of this clone is of an inferred NSI-phenotype, and shows an intact reading-frame over a region sequenced of 1.6kb (chapter 5).

The second construct, A.3, was made by the exchange of a 1.4kb fragment (cassette-3) of the *env* gene vector and has an identical gp120-coding region as clone A.1, which failed to produce infectious virions upon transfection. The *env* gene has been sequenced and shown to contain an intact reading-frame, and a V3 loop sequence of an inferred NSI-phenotype (chapter 5).

Results of Transfections with Constructs P.2 and A.3 (and A.1, E.1 and F.1)

Three transfections with construct P.2 were conducted into COS-1 cells followed by co-cultivation with PBMCs (table 6.5(ii)). In addition, in this set of transfections, the two constructs, E.1 and F.1, previously demonstrated to produce minimally infectious virus, were also transfected. pHXB2-MCS was used as a control. The data from this transfection are given in figure 6.3 and table 6.3.

On day 4 post-transfection, there were p24-positive staining COS-1 cells produced from each transfected culture. However, E.1. and F.1 had fewer p24-positive cells present than the other cultures (table 6.3).

The transfection with pHXB2-MCS showed evidence of virus replication in the first two weeks of culture: supernatant p24 increased to 66ng/ml on day 11, and p24 positive cells could be detected until day 22. However, a drop in the supernatant p24 levels were observed until day 25, after which levels increased again to 70 ng/ml (figure 6.2).

Cultures transfected with E.1 showed early evidence of infection, with p24-positive cells detected up to day 11 post-transfection. p24 levels in the supernatant increased up to day 15 but never reached more than 14 ng/ml, and fell to below 1ng/ml from day 25 onwards.

Each of the transfections with P.2 (P.2a ,-b and -c, in figure 6.2), showed some evidence of viral replication. Within the first week, positive staining PBMCs were observed , although P.2b and -c contained fewer positive cells. Each cloned showed

an increase in supernatant p24 from day 4 to day 8, although the levels in each culture varied widely: P.2a showed an increase of 44ng/ml to 60ng/ml; whereas P.2b showed an increase of <1ng/ml to 4.1ng/ml, and P.2c of 2.8ng/ml to 5.2ng/ml. From day 8 onwards, p24 production from P.2b and P.2c declined steadily with time: supernatant p24 levels dropped to <1ng/ml by day 18, and no p24 positive cells were observed. However, in P.2a there was a dramatic increase in supernatant p24 to 139ng/ml by day 15, despite a small drop to 51ng/ml in the previous time point (day 11). This increase in p24 coincided with the detection of p24 positive cells. However, after day 15, a dramatic decrease in supernatant p24 occurred, reaching 16ng/ml on day 16. Thereafter a gradual decrease was found reaching 4.6ng/ml at day 32. No p24 positive cells could be observed beyond day 18, indicating no viral infection.

F.1 showed a similar pattern of infection as P.2a. Supernatant p24 levels increased steadily from 27ng/ml to 42ng/ml between days 4 and 8, and then increased dramatically to 110ng/ml by day 11. After this, a sharp decrease to 28ng/ml occurred by day 15 and a steady decrease to 3.6ng.ml by day 32 was observed. p24 positive cells could be observed up to day 11, after which no positive cells could be observed.The sharp drop in supernatant p24 levels in P.2a and F.1, coincided with a drop in the p24 level in the supernatant from pHXB2-MCS and a reduction of p24-positive cells at day 22 and no p24 positive cells on day 25. The magnitude of the decrease in p24 production in P.2a and F.1, coincident with a drop in virus production in the control, suggests that culture conditions may have deteriorated at these time points. The fresh PBMCs added to the cultures at time points 11, 15 and 18, may have been less able to support viral replication than the PBMCs used at the beginning of the experiment, and may have diluted the virus. Despite the decline in p24 production, virus from pHXB2-MCS showed an increase in supernatant p24 production from day 24 onwards, and p24-positive staining cells from day 29 onwards.

We attempted to co-culture the viruses obtained from P.2a and F.1 by pooling the supernatant and cells from day 8, 11 and 15 from each culture and co-culturing with fresh, uninfected PBMCs. However, no virus infection could be detected as determined by the absence of p24-antigen staining PBMCs and the inability to detect

Figure 6.3 and Table 6.3 p24-Antigen Production after Transfection into COS-1 Cells followed by Co-Cultivation with PBMCs for Clones pHXB2-MCS, E.1, F.1, and P.2

Figure 6.3 gives the time course of p24-antigen production in culture supernatants
Table 6.3 shows the presence p24-positive staining PBMCs in the transfected cultures.

+ the presence of p24-positive cells at (> 10%)
w+ cultures with less that 10% cells staining for p24 antigen
- cultures with no p24 staining above background

opposite page 202

Figure 6.3

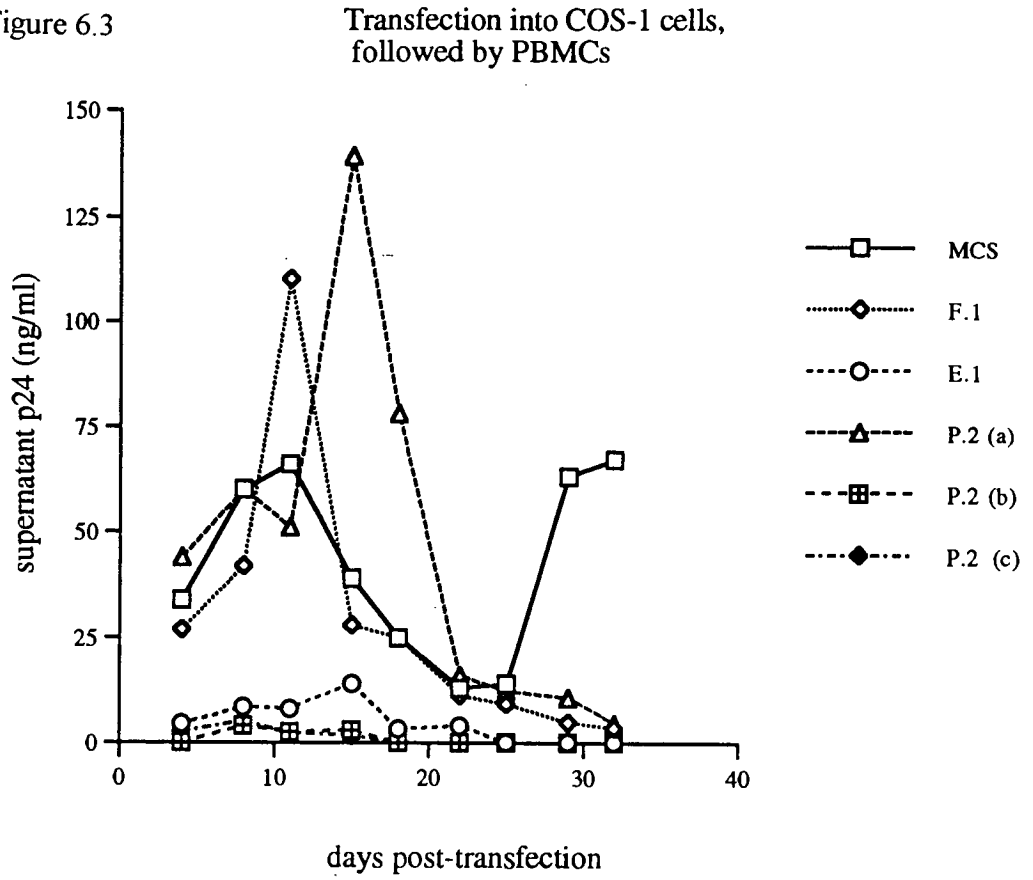


Table 6.3.

	PBMCs days post transfection									
clone	72hr[1]	4	8	11	15	18	22	25	29	32
MCS	+	+	+	+	+	+	w+	-	+	+
E.1	-	W+	+	+	-	-	-	-	-	-
F.1	W+	+	w+	+	-	-	-	-	-	-
P.2a	+	+	+	+	+	W+	-	-	-	-
P.2b	+	-	w+	-	-	-	-	-	-	-
P.2c	+	w+	-	w+	-	-	-	-	-	-

[1] COS-1 monolayer.

The three clones, Q.2, R.2 and S.2, were constructed by exchange of cassette 2 from sequences amplified from the 1989 PBMC DNA sample (chapter 4). We have obtained the V3 sequence from only one of these constructs, R.2, which is predicted to confer an SI phenotype. Although the two remaining clones have not been sequenced, the majority of proviral sequences from 1989 PBMC sample were found to have a D-lineage V3 loop, the majority of which are predicted to confer an SI phenotype (see chapter 5). In consequence, we would expect all three clones to show an SI phenotype.

From the results of the transfections carried out so far, inferred-SI clones would be expected to produce infectious virus from both T-cell line and PBMC cultures. Therefore, two transfections were carried out for each of these constructs, the first into COS-1 cells, followed by co-cultivation with PBMCs, the second directly into the C8166 cells (table 6.5(ii)).

Results of Transfection with Q.2, R.2 and S.2

Viral replication was detected from each of the three constructs, in both PBMCs and C8166 cells, as determined by the production of p24 antigen in the culture supernatant and from the presence of p24-positive staining cells (table 6.5 (ii)). Syncytia could be observed in both PBMCs and C8166 cells transfected with each clone.

Figure 6.4 and table 6.4 show a single experiment following transfection of pHXB2-MCS, Q.2, R.2 and S.2 into C8166 cells. Although the rate of p24 production varied widely between each clone, each could be passaged to a variety of T-cell lines and PBMCs.

A comparison of the infectivity properties of these clones with the two other viable PBMC-derived recombinant viruses, G.1 and H.1, is described further below (section 6.2.6).

Figure 6.4 and Table 6.4 p24-Antigen Production after Transfection into C8166 Cells for Clones pHXB2-MCS, Q.2, R.2 and S.2

Figure 6.4 gives the time course of p24-antigen production in culture supernatants
Table 6.4 shows the presence p24-positive staining cells in the transfected cultures.

+ the presence of p24-positive cells at (> 10%)
w+ cultures with less that 10% cells staining for p24 antigen
- cultures with no p24 staining above background

opposite page 205

Figure 6.4

Transfection into C8166 cells

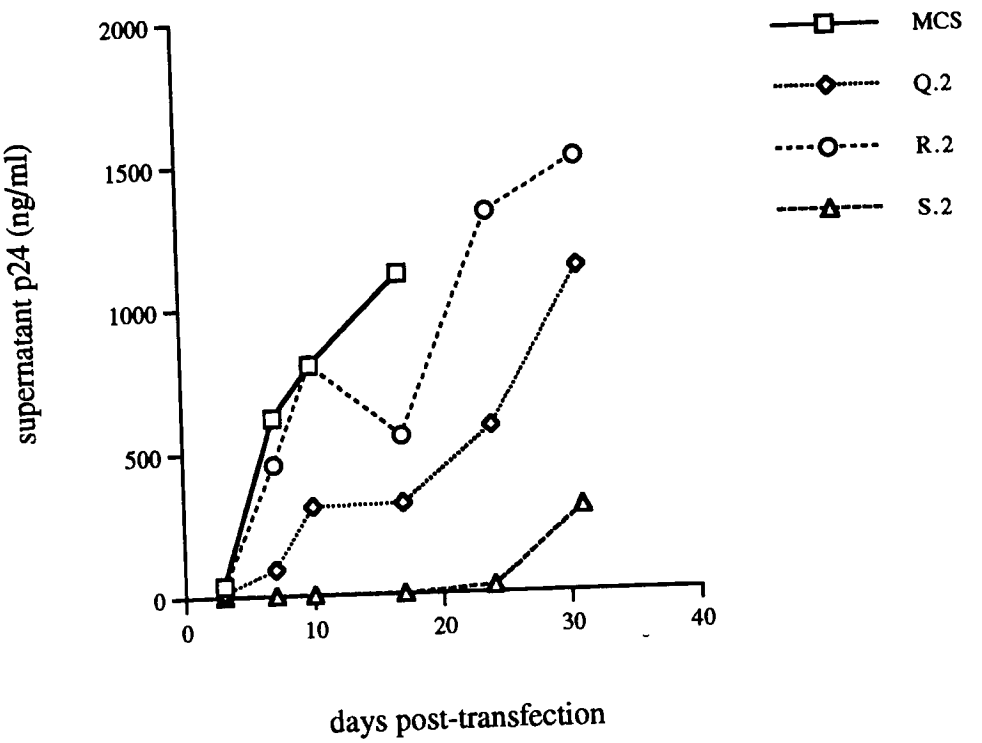


Table 6.4.

	days post transfection						
clone	72hr[1]	3	7	10	17	24	31
MCS	39.0	+	+	+	+	culture stopped	
Q.2	11.0	w+	+	+	+	+	+
R.2	25.0	+	+	+	+	+	+
S.2	<1	-	-	-	-	+	+

[1] p24 (ng/ml) in supernatant .

6.2.5 Recombinant Clones Constructed with $SF2_{MC}$ *env* Sequences

To address further the possibility of a difference in the ability of the PHXB2-MCS expression system to produce infectious virus with NSI, non-T-cell-line tropic *env* genes, compared to SI, T-cell-line tropic *env* genes, we obtained two infectious proviral clones $SF2_{MC}$ and $SF162_{MC}$, which have opposite *in vitro* growth properties (Cheng-Mayer et al, 1989).

Virus from $SF2_{MC}$ is syncytium-inducing and capable of growth in a variety of T-cell-lines, and in PBMCs. In contrast, virus from $SF162_{MC}$ is non-T- cell-line tropic, non-syncytium-inducing, but is capable of growth in both PBMCs and cultures of monocyte-derived macrophages.

The determinants of the growth properties of these two proviral clones have been mapped to the *env* gene. Substitution of a region of the $SF2_{MC}$ *env* gene including the V3 loop and the C2 region into the $SF162_{MC}$ backbone results in a recombinant virus with an acquired ability to grow in T-cell- lines and to produce syncytia, but a loss of ability to grow in cultures of monocyte-derived macrophages. Similarly, the substitution of a region of the $SF162_{MC}$ *env* gene including the V3 loop, and C2 region into the $SF2_{MC}$ clone, confers on the recombinant virus, an ability to grow in cultures of monocyte-derived, but a loss in the ability to grow in T-cell-lines, and an inability to form syncytia (Cheng-Mayer *et al*, 1988)

I therefore aimed to utilise the *env* genes from these two clones to determine the ability of our HXB2 expression-system to produce recombinant virus, expressing the *env* genes known to confer either, an SI-phenotype, or an NSI-phenotype.

However, the cloning sites for the construction of the recombinant viruses of this study, include a *Bst*EII site at the 5' end of the three cassettes (chapter 3). The presence of a *Bst*EII site complicated the construction of the $SF162_{MC}$ recombinant derivative, because it has an additional *Bst*EII site within the *env* gene region to be

exchanged. In consequence, it was not possible under the time constraints of this study to clone the $SF162_{MC}$ *env* gene into PHXB2-MCS.

I have, however, constructed two recombinant viral clones with the *env* gene from $SF2_{MC}$. The first of these clones, HXSF2.1 has the 1.7kb (cassette-1) of $SF2_{MC}$, incorporated into PHXB2-MCS. The second, HXSF2.2, incorporates the 2.5kb (cassette-2), into PHXB2-MCS (chapter 4).

These two constructs differ, therefore, only in the length of the $SF2_{MC}$ fragment incorporated. The presence of the SF2 *env* gene in HXSF2.1 and HXSF2.2 should confer upon these viruses a T-cell-line tropic, syncytium-inducing phenotype. We therefore carried out transfections with these clones into both PBMCs and the cell lines H9, C8166 and SupT1. In addition, we carried out a transfection with the two parental clones $SF2_{MC}$ and $SF162_{MC}$ with PBMCs (table 6.4(iv)).

Results of Transfections with $SF2_{MC}$, $SF162_{MC}$, HXSF2.1 and HXSF2.2

Transfection of both parental clones, $SF2_{MC}$ and $SF162_{MC}$ into COS-1 followed by co-cultivation with PBMCs resulted in the production of infectious virus. There was, however, a significant delay in the production of p24 antigen with both clones, compared to p24-production from PHXB2-MCS. With $SF2_{MC}$, p24 could be detected in the supernatants of the transfected COS-1 cells, and remained at detectable but low levels in the culture supernatant until day 5. After day 5, no p24 could be detected until day 20, when high levels of p24 (>1000ng/ml) could be detected in the supernatant and p24-positive cells could be observed. With $SF162_{MC}$, there were no signs of p24 production until day 19 post-transfection when p24-positive cells could be observed. At day 23 and onwards, low levels (13ng/ml) of p24 could be detected in the supernatant. DNA sequence analysis confirmed the identity of the virus produced on transfection (data not shown).

Virus from the $SF2_{MC}$ cultures was readily transmitted to T-cell line H9 by co-culture.

High levels of p24 (>1000ng/ml) were detected in the supernatant at day 10 post-co-cultivation, and p24-positive cells were observed. Virus from $SF162_{MC}$ cultures could not be transmitted to H9 T-cell line by co-culture; no 24 positive cells were observed and only low levels of p24 were detected, probably reflecting carry-over of virus particles and/or infected PBMCs from the co-culturing. The inability to transmit $SF162_{MC}$ to H9 cells, is in agreement with the non-T-cell-line tropic phenotype found previously with this clone (Cheng-Mayer *et al*, 1989).

In contrast, from five transfections with HXSF2.2 and a single transfection with HXSF.1, no infectious virus could be recovered; no p24-positive cells could be detected beyond day 4 post-transfection, and the levels of p24 antigen in the supernatant declined to below 1ng/ml after one week of culture (data not shown).

Lack of Infectivity with HXSF2.1 and HXSF2.1

The failure to produce viable viruses from HXSF2.1 or HXSF2.2, could be due to either a failure in the interaction between $SF2_{MC}$ *env* sequences and HXB2 background, or due to the introduction of errors during amplification and cloning. The recovery of infectious virus from clone $SF162_{MC}$ demonstrates that our PBMC culture-system can successfully propagate virus of an NSI-phenotype.

Table 6.5 Summary of the Transfections of HIV-1 Recombinant Proviral Clones

[a] SI: syncytium-inducing, NSI: non-syncytium-inducing

opposite page 209

i) PBMC-derived constructs: cassette-1 with *Pfu*-amplified *env* sequences:

clone	predicted phenotype[a]	cell type	No. attempts	No. attempts positive
A.1	NSI	C8166 SupT1 U937 COS:PBMC	5 1 1 2	0 0 0 0
B.1	NSI	COS:PBMC COS:U937	1 1	0 0
C.1	NSI	C8166 H9 PBMC COS:PBMC COS:U937	1 1 2 1 1	0 0 0 0 0
D.1	NSI	COS:PBMC COS:U937	1 1	0 0
E.1	NSI	COS:PBMC COS:SupT1	3 1	2? 0
F.1	NSI	COS:PBMC COS:SupT1	3 1	2? 0
G.1	SI	H9 SupT1 COS:PBMC COS:H9	1 1 2 1	1 1 2 1
H.1	SI	C8166 H9 PBMC COS:PBMC COS:H9	7 2 3 2 1	6 1 0 2 1
I.1	NSI	C8166 H9 Jurkat$_{tatIII}$ PBMC	1 1 1 3	0 0 0 0
J.1	NSI	Jurkat$_{tatIII}$ COS:SupT1 COS:PBMC	1 1 0	0 0 0

Table 6.5 continued

with *Taq*-amplified *env* sequences:

clone	predicted phenotype[a]	cell type	No. attempts	No. attempts positive
K.1	SI	COS:PBMC COS:H9	1 2	0 0
L.1	SI	C8166 COS:PBMC	2 1	0 0
M.1	SI	C8166 COS:PBMC	2 1	0 0
N.1	SI	H9 COS:PBMC COS:H9	1 1 1	0 0 0

ii) PBMC-derived constructs: cassette-2

clone	predicted phenotype[a]	cell type	No. attempts	No. attempts positive
P.2	NSI	COS:PBMC	3	3?
Q.2	SI?	COS:PBMC C8166	1 1	1 1
R.2	SI	COS:PBMC C8166	1 1	1 1
S.2	SI?	COS:PBMC C8166	1 1	1 1

iii) PBMC-derived constructs: cassette-3

clone	predicted phenotype[a]	cell type	No. attempts	No. attempts positive
A.3	NSI	COS:PBMC	1	0

Table 6.5 continued

iv) Control Constructs

clone	predicted phenotype[a]	cell type	No. attempts	No.attempts positive
pHXB2-SPT	SI	C8166	6	3
		Jurkat$_{tatIII}$	1	1
		PBMC	1	0
		COS:PBMC	1	1
pHXB2-MCS	SI	C8166	4	4
		H9	2	2
		Jurkat$_{tatIII}$	2	2
		SupT1	1	1
		COS:C8166	2	2
		COS:H9	2	1
		PBMC	4	1
		COS:PBMC	6	5
pHXB2-MCSΔ*env*	non-infectious	C8166	2	0
		H9	1	0
		COS:H9	1	0
HXSF2.1	SI	COS:C8166	1	0
HXSF2.2	SI	COS:PBMC	1	0
		COS:H9	1	0
		H9	1	0
		COS:SupT1	1	0
		COS:C8166	1	0
$SF2_{MC}$	SI	COS:PBMC	1	1
$SF162_{MC}$	NSI	COS:PBMC	1	1

6.2.6. *In Vitro* Growth Properties of the Recombinant Viruses

The following sections describe the results of work carried out to determine in more detail the biological properties of the five viable PBMC-derived proviral clones, G.1, H.1, Q.2, R.2, and S.2. The virus propagation was carried out entirely by Dr Robert Walker and Mrs Elizabeth Harvey.

Prior to determining in more detail the biological properties of the infectious clone *in vitro*, I carried out a sequence determination of two of the clones G.1 and H.1, to examine whether any sequence changes had occurred during the propagation of the clones *in vitro*. After a prolonged period of culture of 6 weeks in C8166 cells, total DNA was extracted from the cultures of G.1 and H.1. and PCR-amplified in the V3 to V5 region (chapter 2). The sequence of the V3 to V5 region comprising of 1kb was obtained. No differences could be detected between the sequence obtained after 6 weeks of culture and the sequence of the original proviral variant (data not shown).

The Cellular Host-Range and Syncytium-Inducing Ability

The cellular host-range of the five infectious PBMC-derived recombinant proviral clones (G.1, H.1, P.2, Q.2 and R.2) was analysed by infecting transformed cell lines with cell-free virus derived from the clones, or by co-cultivation with virus-infected cells recovered from the transfection with the proviral DNA.

Each clone was capable of productively infecting T-cell lines C8188, H9, SupT1, and MT2. In addition, each clone induced the production of syncytia in the cell lines tested.

Table 6.6 shows the maximum p24 antigen levels produced with each clone following

growth in cell lines H9, C8166 and PBMCs. In H9 and C8166 cells each clone analysed produced high levels of p24 in the supernatant, ranging from 700ng/ml to over 1000ng/ml. These values are comparable to the controls isolates HIV-1_{IIIB} and HIV-1_{SF2}, although HIV-1_{SF2} reached just less than 500ng/ml in C8166 cells.

By comparison, each viral construct grew to lower levels in PBMCs than T-cell lines, with the exception of clone H.1 which reached levels of >1000ng/ml. There was considerable variation in the levels of p24 produced in infections with the same construct but where different batches of PBMCs were used. For example, H.1 showed a maximum p24 level of 67ng/ml, despite producing >1000ng/ml with different PBMCs. Similarly, R.2 showed 55ng/ml on one occasion, and 212ng/ml with a different batch of PBMCs. PBMCs are generally less able to support HIV-1 replication, which in part, reflects the less robust nature of PBMCs and the lower proportion of CD4-positive cells in PBMC cultures compared to T-cell lines.

Virus Infectivity

One of the earliest observations made with the PBMC-derived clones, was that although each produced similar high levels of p24 in the culture supernatants, the infectivity titres of cell-free virus was between two or three orders of magnitude lower than the parental virus, pHXB2-MCS and the isolate HIV-1_{IIIB}. In contrast, virus transmission by co-cultivation appeared to be equally efficient with the PBMC-derived recombinant virus and pHXB2-MCS and HIV-1_{IIIB}; in each case, virus spread and amplification, detected as the presence of syncytia throughout the cultures, was readily observed two or three days post-co-cultivation.

To investigate this further, infectivity titres were calculated for both cell-free and cell-to-cell transmission (table 6.7). Cell-free infectious titres from the two clones constructed by the exchange of cassette-1, H.1 and G.1, were $1.0x10^2$ and $8.3x10^2$, respectively. The infectious titres for the three clones constructed through exchange of cassette-2, were 10-fold higher for Q.2 at $3.6x10^3$, and greater than 100-fold higher

for R.2 and S.2, at $8.0x10^4$ and $2.0x10^4$, respectively.

The infectivity titres from the PBMC-derived constructs were between 10- and 1000-fold lower than the control isolate HIV-1_{IIIB}, and pHXB2-MCS, which were $1.3x10^6$, and greater than 10^5, respectively. The isolate HIV-$1S_{F2}$, in our hands, produced a relatively low infectious titre of $2.5x10^3$.

Infectious titres determined from cell-to-cell infection were in the range of $1x10^5$ and $1x10^4$ for the PBMC-derived clones. These were comparable to the infectious titres determined from the control constructs HIV-1_{IIIB} and pHXB2-MCS and HIV-1_{SF2}.

The low infectivity titres of cell-free virus obtained from the PBMC-derived constructs have precluded their use in neutralisation assays, which require large amounts of virus at high infectivity titre.

Table 6.6. Maximum p24 levels (ng/ml)

	Maximum p24 levels (ng/ml)		
	cell type		
virus	C8166	H9	PBMCs
pHXB2.MCS	1446	1224	54
G.1	1014	>1000	440
H.1	1079	>1000	67, 1160
Q.2	1125	1119	27
R.2	1510	712	55, 212
S.2	n.d.	881	33
HIV-1_{SF2}	493	>1000	>1000
HIV-1_{IIIB}	856	1430	n.d.

Table 6.7. Virus infectivity titres

	Virus infectivity titres		
virus	$TCID_{50}$/ml	$TCID_{50}/10^6$cells	p24 (ng/ml)
pHXB2.MCS	$9.3X10^5$ $4.0x10^5$	$3.2x10^5$ $2.5x10^5$	865 736
G.1	$8.3x10^2$	$2.5x10^5$	380
H.1	$1.0x10^2$	$7.0x10^4$	234
Q.2	$3.6x10^3$	$4.0x10^4$	1119
R.2	$8.0x10^4$	$3.2x10^5$	712
S.2	$2.0x10^4$	$4.4x10^4$	889
HIV-1_{SF2}	$2.5x10^3$	$1.4x10^5$	316
HIV-1_{IIIB}	$1.3x10^6$	$4.0x10^5$	1430

Chapter 7
Discussion

7.1. The Immune Response to HIV-1 Infection

One of the most important questions concerning the nature of HIV-infection, is how HIV maintains a chronic infection despite the presence of a strong antiviral immune response.

There is evidence that the immune system can control virus replication following primary infection when high titres of infectious virus are found. A rapid decrease in plasma viraemia occurs within several days to weeks, concomitant with the first detectable immune response (Daar *et al*, 1991; Clark *et al*, 1991). Cytotoxic T-lymphocyte (CTL) responses are detected first, before neutralising antibodies (Koup *et al*, 1994; Arendrup *et al*, 1992; Albert *et al*, 1990; Ariyosha *et al*, 1992; Moore *et al*, 1994b). Moore *et al*, 1994b, have noted that although the CTL response occurs first, and will kill infected cells, neutralising antibodies are probably responsible for the rapid clearance of infectious virus from the plasma at this time point.

A CTL response to HIV-1 can be detected through out the asymptomatic period (Clerichi, 1993). However, in infected individuals there is frequently found a loss of function of the T-helper cell subset associated with augmenting the cell-mediated immunity (the Th_1 subset) and a concomitant activation of the Th_2 subset, which stimulate humoral immunity and the production of antibodies. This imbalance will almost certainly lead to a reduced function of cellular immunity.

Long term survivors, who show a normal and stable CD4+ T-cell count after 12 to 15 years of HIV-1 infection, show a strong CTL response and a broad neutralising antibody response. This vigorous immune response may be responsible for a low viral load and a good clinical prognosis (Cao *et al*, 1995). However, other factors may be important: the CD4 cells on long term survivors may be less susceptible to infection, or that they harbour attenuated viruses (Zhang *et al*, 1996; Iverson *et al*, 1995; Cocchi *et al*, 1995).

Moore and Ho, 1995, have postulated that ultimately, the immune response is unable

to prevent disease progression, because antibodies are ineffectual against HIV-1 in the lymph nodes and lymphoid tissues, where a large proportion of HIV-1 is harboured.Two mechanisms for this have been proposed: firstly, the high cell density in the lymphoid tissues reduces the effectiveness of antibodies; secondly, the spread of HIV-1 may occur principally by cell-to-cell infection in lymphoid tissues, to which antibodies are less efficient at blocking.

7.2 Antigenic Variation in Infectious Diseases

Another factor, which forms the basis of the work presented in this study, is that HIV maintains a chronic infection by continually producing novel antigenic variants which are not recognised by the contemporaneous immune response. In this way, the virus will be able to 'escape' destruction by the host's immune system. This phenomenon is known as antigenic drift, and has been proposed for chronic infections with other lentiviruses (such as Visna virus of sheep and equine infectious anaemia virus of horses; Thormar *et al*, 1983; Lutley *et al*, 1983), and in a number of bacterial and protozoan infections, such as *Borrelia recurrentis*, African trypanosomes and the malaria parasite (Bloom, 1979; Mims, 1986). In these cases the production of antigenic variants during the course of a single infection is an important factor promoting their persistence in the body.

With a number of viral infections, antigenic variation does not occur during the course of a single infection, but may occur outside the host, and so re-infection with a novel antigenic variant can occur at a later stage. This is observed for influenza A and influenza B viruses, foot and mouth disease virus, and human rhinoviruses (Mims, 1982; White and Fenner, 1994). Perhaps the most striking example is that of type A influenza virus - in which antigenic variation is responsible for the sporadic occurrence of pandemics of influenza within the human population. There are two viral envelope proteins haemagglutonin (HA) and neuraminidase (NA), both of which are highly glycosylated. Type A influenza virus may be classified into subtypes according to the antigenic characteristics of the surface antigens HA and NA. The surface antigens undergo two types of antigenic variation: *antigenic drift* which result from mutation in

either the NA or HA genes, which are no longer recognised by the immune system; and *antigenic shift* which results from the acquisition of a novel HA or NA gene, producing a new human subtype, which is antigenically distinct. This is thought to arise by recombination between different strains of influenza A virus leading to a new combination of NA or HA subtypes. It is also thought that a new subtype may arise in the human population through acquisition of an avian subtype NA or HA, by recombination of avian and human influenza type A strains, facilitated by the segmented RNA genome (Webster *et al*, 1982). Antigenic shift is believed to be responsible for pandemic outbreaks of influenza that have occurred in 1918, 1946, 1957 (Asian 'flu), 1968 (Hong Kong 'flu).

Viruses which evolve rapidly and produce novel antigenic variants within the infected host population, or within an individual host, pose problems for vaccination. This is principally because vaccination to one strain may have no effect against another. Much research has been vested into quantitating the number of antigenic variants in a given geographical area, and in predicting which variants will arise in the future by studying the patterns and processes of viral evolution (Webster *et al*, 1982; Louwagie *et al*, 1993; Holmes *et al*, 1992; Moore *et al*, 1996).

7.3. The Genetic Variation of HIV-1

HIV, in common with other viral species and especially RNA viruses, exists within the infected individual as a heterogeneous population of variants. This distribution of variants has been termed a 'quasispecies' and is the direct result of a high mutation rate and a large population size. Competition between the variants under different environmental conditions shapes the distribution. The unusual feature of the quasispecies is that the fitness of particular sequences is not just a function of the sequence *per se* but is also related to the mutational distance between them and the fittest variants (Eigen *et al*, 1988; Domingo *et al*, 1993; Nowak, 1992).

This distribution of variants is thought to allow HIV to adapt quickly to new environments, such as those formed by the host's evolving immune response where antigenic variants may be selected (Albert *et al*, 1990), or following the administration

of drug therapy where drug resistant forms of the virus may be selected (Larder *et al*, 1989), or in the adaptation of HIV-1 to a particular cell or tissue type, where *env* mutations confer different cellular host-ranges (Korber *et al*, 1994; Cheng-Mayer *et al*, 1989).

7.4 The Principal Targets of Neutralising Antibodies elicited following HIV-1 Infection

For laboratory strains, the principal target of neutralising antibodies found in natural infection with HIV-1 is an epitope in the third hypervariable region of the surface protein gp120, which forms a stem-loop structure - 'the V3 loop' (figure 1.5, chapter 1) (Rusche *et al*, 1988; Javaherian *et al*, 1989; Zwart *et al*, 1994. A second important target of neutralising antibodies is a complex epitope which makes up the binding-site for the cellular receptor molecule, CD4 (Chamat *et al*, 1992; Kang *et al*, 1991). The CD4 binding-site epitope is formed from residues in the conserved regions C2, C3 and C4 of gp120 (Lasky *et al*, 1987; Olshevsky *et al*, 1990), and interacts with the V3 loop (Wyatt *et al*, 1992; Moore *et al*, 1993a).

7.4.1 The Function of the V3 loop

The V3 loop as well as being defined as the principal neutralising determinant of HIV-1, is also an important determinant of the *in vitro* viral phenotype: the cellular host-range, replication-rate and syncytium-inducing ability (Hwang *et al*, 1991; De Jong *et al*, 1993a, 1993b; Stamatatos *et al*, 1994). In general, HIV-1 isolates fall into two groups: slower replicating variants that can grow in macrophages and CD4 T-lymphocytes, but are non-syncytium-inducing (NSI) and are unable to replicate in T cell-lines (Schuitemaker *et al*, 1992; Connor *et al*, 1993b); the other group consists of faster replicating variants which can grow in macrophages, T-lymphocytes and T-cell lines, these variants can usually induce the formation of syncytia (syncytium-inducing or SI) (Connor *et al*, 1993b; Tersmette *et al*, 1989a and 1989b). It is the amino acid sequence and charge of the V3 loop which determines these phenotypes: a low net charge and a limited sequence diversity from the subtype B consensus sequence predicts a non-syncytium-inducing phenotype and a non-T-cell-line tropic phenotype; a high net

charge and a high sequence diversity from the subtype B consensus sequence confer a syncytium-inducing phenotype and an ability to replicate in T-cell-lines (Milich *et al*, 1993; Fouchier *et al*, 1992; Chesebro *et al*, 1992).

There is considerable natural variation in the amino acid sequence of the V3 loop both within an individuals and between individuals (Simmonds *et al*, 1991; Holmes *et al*, 1992; Wolfs *et al*, 1991; Fouchier *et al*, 1992; Cichutek *et al*, 1992; Leigh Brown and Holmes, 1994; Leigh Brown, 1991). Many of the variable sites in the V3 loop alter antibody binding to synthetic peptides representing the V3 epitope (Wolfs *et al*, 1991; Moore *et al*, 1995), and can give rise to variants with different sensitivities to neutralisation by serum antibodies (Moore *et al*, 1996; Looney *et al*, 1988). *In vitro* studies with HIV-1 clones have shown that resistant forms can arise *de novo* under selection from neutralising antibodies to both the V3 loop and the CD4 binding-site epitope (McKeating *et al*, 1989, 1993; Yoshiyama *et al*, 1994; Masuda *et al*, 1990; Robert-Guroff *et al*, 1988). With the V3 loop, single amino acid substitutions within the loop, or at sites distant from the loop have been shown to confer resistance (McKeating *et al*, 1989; Wahlberg *et al*, 1991).

However, as described above, the variation in the V3 loop also influences the *in vitro* viral phenotype: early on in infection viral isolates tend to be of non-syncytium-inducing phenotype, whereas later in infection faster-replicating, syncytium-inducing variants also emerge (Connor *et al*, 1993b and 1994; Schuitemaker *et al*, 1992). These phenotypic changes correlate with amino acid changes in the V3 loop (Fouchier *et al*, 1992; Connor *et al*, 1993a), and may be one of the forces of selection acting upon the virus population *in vivo*.

7.4.2 Characterising the Evolution of the V3 Loop

Despite extensive studies of sequence variation, we are still uncertain of the role of selection in driving the evolution of the HIV-1 *env* gene, and in particular the V3 region (Leigh Brown and Holmes, 1994; Domingo *et al*, 1993; Moore and Ho, 1995; Donaldson *et al*, 1994b).

the proportion of synonymous (K_s) versus nonsynonymous (K_a) nucleotide substitutions within the V3 region over a 7-year period from a single individual (patient 82 of this study). In year 3 post-infection, there was a ten fold higher mean number of nonsynonymous substitutions (K_a), which is indicative of positive selection for amino acid change. By year 5, the ratio of K_s to K_a, had increased to greater than 1, indicating weak negative selection against amino acid changes. This increase in K_s/K_a correlated with a low CD4 count, is consistent with a role for the immune system in selecting variants.

However, other investigations have failed to show consistent K_s/K_a ratios in the V3 region of less than 1, suggesting a random element in the sequence evolution (Zhang *et al*, 1996). Similarly, a variable K_s/K_a ratio (of between 0.2 to 2) was found by Lukashov et al, 1995, in an extensive study of 44 individuals. In each of these studies, a significant difference in the rates of evolution was found between the patients. This is consistent with specific host factors, such as differing immune responses, in driving the evolution.

Leigh Brown (1996) has applied the neutral theory to data from patient 82, to determine whether the pattern of variation in the V3 region fits that expected for selectively neutral sequences, whose evolution will be governed by mutation rate and random drift, or whether the pattern of variation is significantly different from that of selective neutrality. If so, this would indicate that selection was driving the evolution of the V3 region. At only two time points out of six (year 4 and year 6) did amino acid sequences show a level of diversity greater than expected from neutrality, suggesting a role for selection. Further analysis showed that only the V3 loop itself showed a significant departure from neutrality, the regions flanking the V3 loop did not show a departure from neutrality. From this analysis, only the V3 loop is under selection in this patient. Interestingly, the diversity in the V3 loop in years 4 and 6 from patient 82 is at sites important in determining cell tropism, which may be the selective force acting upon the V3 loop. (This was described in chapter 5, section 5.5).

So we have some evidence that selection is acting upon the V3 loop; selection may be from the immune response, or for variants with a different cellular host-range. It may

from the immune response, or for variants with a different cellular host-range. It may be difficult to determine which is the most important, as the same amino acid changes can alter both the neutralising sensitivity and the cell tropism.

However, other factors aside from selection, may be important in shaping the evolution of HIV-1: the clonal activation of individual T-cells harbouring HIV-1, or random drift of selectively neutral variants (Domingo *et al*, 1993). Indeed, Leigh Brown (1996) has estimated that the effective population size (N_e) of HIV-1 in the host is sufficiently small that random drift may have a significant effect upon viral evolution.

7.5. A Model of Antigenic Diversity in HIV-1 Pathogenesis

The extensive variation displayed by HIV-1 *in vivo* has been the basis of mathematical model: the 'antigenic diversity threshold' model which examines the role of antigenic variation upon HIV-1 pathogenesis (Nowak *et al*, 1990).

During the early stages of HIV infection, prior to the development of the immune response, there will be an outgrowth of the fastest-replicating variants, reducing the observed variation of the population. A reduction in variation in the *env* gene upon primary infection has been demonstrated by a number of groups (Zhu *et al*, 1993; Zhang *et al*, 1993; Delwart *et al*, 1993).

Following seroconversion, the model states that the immune response will select for antigenic variants that have escaped neutralisation. With time, this process will give rise to virus diversification and variation will accumulate in epitopes recognised by the immune response. Diversification in the *env* gene has been observed by a number of groups (Holmes *et al*, 1992; Wolfs *et al*, 1990 and 1992; McNearney *et al*, 1992; Kuiken *et al*, 1993), but the role of immune selection remains undetermined (see above).

With time, there will be an increase in antigenic diversity which the immune system will find more and more difficult to control. The model proposes that an imbalance between immunological and viral diversity will develop, on the assumption that each

The antigenic diversity will increase until it reaches a certain level, or a 'diversity threshold', at which there are too many variants for the immune system to contain. Once this threshold is reached the immune system will collapse and the patient will progress to AIDS. Viral replication in AIDS will escalate due to the absence of immune control. In the absence of selection for antigenic variants, the fastest replicating variants will predominate and this will reduce the variation observed. A low sequence diversity in AIDS has been observed by some groups (McNearney *et al*, 1990) but not by others (Donaldson *et al*, 1994b).

The degree of genetic diversity has been examined with respect to the rate of disease progression. Lukashov *et al*, 1995, observed that in a large data set of 44 individuals, non-progressors accumulated a greater number of amino acid changes in the V3 loop, than those who progressed rapidly to AIDS. In a second study, Delwart *et al*, 1994, found that long-term infected asymptomatics showed the highest levels of genetic complexity in the *env* gene. A rapid CD4-cell decline and progression to AIDS was often associated with lower complexity. It may be that a strong immune response is present in the non-progressors and this may drive *env* and V3 sequence evolution, leading to a greater genetic complexity. Individuals progressing rapidly to AIDS may have a weaker immune response, thus weakening the selective pressure for antigenic diversification.

However, in contrast to this, a study of two long-term non-progressors, (who showed a normal and stable CD4 count over a 5 years of infection with HIV-1), showed each to have a very low viral sequence diversity and viral load throughout this period of asymptomatic infection (Haigwood *et al*, 1993). The low diversity is presumably due to immune control of viral replication, which also results in a very low viral load.

Thus, it may be argued that a strong immune response will lead to a low viral load and a low viral diversity, and a good disease prognosis, as was seen by Haigwood *et al*, 1993 and Strunnikova *et al*, 1995. However, it may also be argued that the extensive diversity seen in slower-progressors is due to a strong immune response, which selects for escape mutants - as was seen by Lukashov *et al*, 1995. However, this is in contrast to the predictions of the antigenic diversity threshold model. According to the model

the antigenic sites of HIV-1 evolve more rapidly in individuals progressing rapidly to AIDS, as the antigenic diversity threshold will be reached quicker. In contrast, in individuals who do not progress, who remain asymptomatic for longer, evolution of antigenic sites should be slower, and so it will take longer to reach the antigenic diversity threshold.

7.6. Characterising the Immune Response to the V3 Loop using Synthetic Peptides

The antibody response to HIV infection has been extensively examined by determining the reactivity of sera to *in vitro* synthesised peptides representing specific epitopes of gp120 or gp41. This approach has been widely used to characterise the antibody responses to the V3 loop. For example, this method has been used in the characterisation of virus isolates found in different geographical locations to determine the different V3 serotypes (Moore *et al*, 1994c). Peptide studies have also been used to investigate HIV transmission, for example, by mother-to-child transmission, or through sexual contact, with the aim of determining the correlates of transmission and non-transmission (Fiore *et al*, 1994). Peptides have also been used to analyse the humoral immune response during the course of natural infection - in cross-sectional and longitudinal studies, or to study the humoral immune response following vaccination of humans and animals (Wahlberg *et al*, 1991; Wolfs *et al*, 1991; Zwart *et al*, 1992).

There are a number of draw-backs to the use of peptides for characterising the immune responses to HIV-1 infection. Firstly, V3 peptides do not measure the neutralising antibody response to the V3 loop, but measure the binding of serum antibodies to the peptides and these may not have a neutralising property. Secondly, it has been shown that V3 peptides do not reproduce the antigenicity of the V3 loop when present on the native envelope protein complex (Moore, 1993). Although the V3 loop is a linear epitope (one which requires the correct linear amino acid sequence to bind antibody and may be mimicked by short peptide sequences), the binding of antibodies is influenced by the conformation of the loop, which may vary according to the sequence and conformation of the rest of the protein molecule (McKeating *et al*, 1989; Klasse *et al*,

1993; Haigwood *et al*, 1993). Ideally, the envelope protein should be assayed as an oligomeric complex, not as gp120 monomers, as many of the antigenic sites are more exposed on gp120 monomers than on the native envelope protein complex, which consist of oligomers of gp120 in association with gp41 (Moore *et al*, 1995; Sullivan *et al*, 1995). Indeed, to express and assay the native envelope protein complex of variants of HIV-1 found *in vivo*, was the principal aim of the work described in this thesis.

Studies with V3 peptides have, however, shown a number of important results about the antibody-specificity during the course of natural infection. Antibodies to the V3 loop elicited early in infection have been shown to be type-specific - that is will only react with a limited number of peptides of different sequence (Zwart *et al*, 1992). Wolfs *et al*, 1991, have shown that in one individual the V3 sequences that emerge during the course of infection were accompanied by the emergence of a new population of V3-specific antibodies. In a longitudinal study of forty-six seroconverters, Zwart *et al*, 1992, found that in the majority of patients the specificity of antibody response to V3 peptides did not change during the course of infection. In a few patients, the later antibody response had a broader activity. Only three patients developed a completely new serum specificity.

V3 peptides have been used to adsorb anti-V3 antibodies present in HIV-1-positive sera, which can then be eluted and used in neutralisation assays. These types of studies have shown that a large proportion of neutralising activity in sera is directed to the V3 loop (Zwart *et al*, 1994; Profy *et al*, 1990; Wolfs *et al*, 1991). These antibodies are type-specific and will only neutralise a limited number of strains.

7.7 Examination of the Neutralising Antibody Response to HIV-1 Infection

7.7.1 Differing sensitivities to neutralising antibodies of primary isolates and laboratory strains

Up until recently, the majority of neutralisation studies have been carried out with laboratory-adapted strains such as HIV-1_{NY5}, HIV-1_{IIIB}, HIV-1_{MN} and have used T-cell

lines to passage the virus and conduct the neutralisation assays (Zwart *et al*, 1994; McKeating *et al*, 1989). However, more recent studies have shown that there are significant differences between the *in vitro* properties of laboratory-adapted strains and those of primary isolates, particularly in their sensitivities to neutralising antibodies (Moore *et al*, 1995; Sullivan *et al*, 1995; Wrin *et al*, 1995; Bou-Habib *et al*, 1994).

Although there is not a strict definition of primary isolates, they are those which have been obtained by limited cultivation of patient PBMCs, or plasma, with uninfected PBMCs (Fenyo *et al*, 1989; Tersmette *et al*, 1988; Schuitemaker *et al*, 1991; Gao *et al*, 1996; Ho *et al*, 1989). In contrast, laboratory isolates are those that have been extensively passaged in human T-lymphoid cell lines, such as the 'prototype' isolates HIV-1_{IIIB}, HIV-1_{SF2}, HIV-1_{MN}, HIV-1_{SF162} (Shaw *et al*, 1984; Cheng-Mayer *et al*, 1991). The majority of primary isolates are non-syncytium-inducing and replicate in T-lymphocytes and monocyte-derived macrophages. However, most primary isolates do not grow well in cell lines; those that can be passaged in T-cell lines, generally show a syncytium-inducing phenotype (Fouchier *et al*, 1992; Koot *et al*, 1992). In contrast, and by definition, laboratory-adapted isolates generally grow well in a variety of T-cell lines and are generally of a syncytium-inducing phenotype.

Moore *et al*, 1995b, found that primary isolates are more resistant to neutralisation by HIV-1 anti-sera, and certain monoclonal antibodies (MAbs). This is similar to earlier findings showing that laboratory-adapted isolates show a 100 to 1000 fold increased sensitivity to neutralisation by sCD4, a property that correlates with increased shedding of gp120 on exposure to sCD4 (Daar *et al*, 1990). By contrast, primary isolates are relatively resistant to neutralisation by sCD4, which in part, explains the inefficacy of the administration of sCD4 to patients as an anti-viral agent.

The relative resistance of primary isolates to antibodies has been confirmed by the work of Sullivan *et al*, 1995. In this study, the envelope protein complexes of three primary isolates were more resistant to monoclonal antibodies to the CD4 binding-site and the V3 loop, than the envelope complexes from two laboratory-adapted strains. In addition, the primary isolates had a higher density of envelope glycoprotein spikes on the virions, which may contribute to their increased resistance to neutralising antibodies.

Further to this, the adaption of a primary isolate to growth in a T-cell line, which involved sequence changes in the V2, C2 and V3 regions of the surface protein gp120, also increased the susceptibility to neutralisation by V3-specific MAbs (Wrin *et al*, 1995). In another study, it was found that the V3 neutralising-epitope on the primary isolate HIV-$1_{JR\text{-}CSF}$ is presented in a form where it is less-accessible to V3-directed antibodies. The V3 loop epitope was said to be of a 'cryptic' nature. Adaptation of HIV-$1_{JR\text{-}CSF}$ to a T-cell line resulted in amino acid changes to the V3 region, which increased the net charge of the loop, and rendered the virus more susceptible to neutralisation by V3 monoclonal antibodies (Bou-Habib *et al*, 1994).

7.7.2 A Dual Configuration Proposed for the HIV-1 Envelope Protein

Based upon this evidence, Moore and Ho (1995) have proposed that the envelope protein complex of primary isolates has a configuration that is fundamentally different to that of laboratory-adapted strains. They suggest that the configuration of the envelope complex of primary isolates is such that it shields the major neutralising epitopes, especially the V3 loop and the CD4 binding-site, from access to neutralising antibodies. The CD4 binding-site of primary isolates is described as being predominantly in a 'closed' configuration (figure 7.1). In contrast, the envelope protein complex of laboratory-adapted isolates will have evolved during their extensive propagation *in vitro*, to a configuration which enhances their *in vitro* replicative ability but renders them more susceptible to neutralisation by sCD4, monoclonal antibodies and HIV-1 positive sera. In this case, the CD4 binding-site is described as being predominantly in an 'open' configuration.

Specifically, Moore and Ho, 1995, have predicted that the amino acid changes producing a T cell-line adapted virus, alter the association between the V3 loop and residues in C3 and C4 which contribute to the CD4 binding-site. These changes are believed to move the V3 loop into a more exposed position and uncover the CD4-binding site. Amino acid changes in the V3 loop, and elsewhere in gp120 or gp41, have been shown to exert an effect on viral phenotype (Andeweg *et al*, 1993; Boyd *et al*,

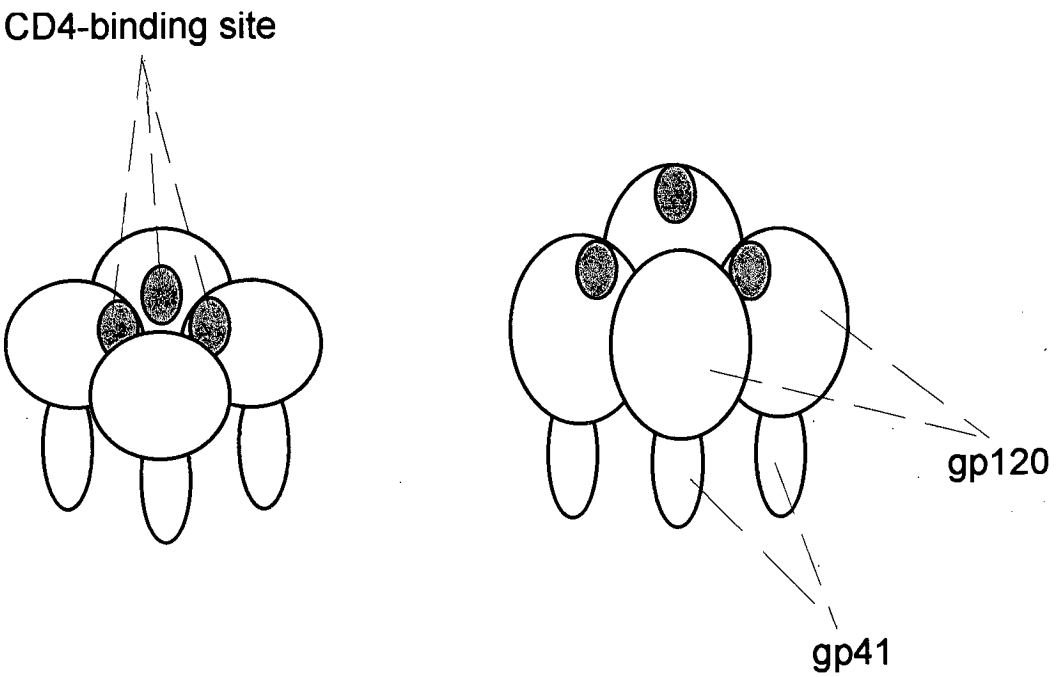
Figure 7.1. Schematic Representation of the Change in Configuration of the Envelope Proteins of HIV-1 following Adaptation of Primary Isolates to growth in T-Cell Lines.

This figure is adapted from Moore and Ho, 1995.

opposite page 229

Primary isolates

Cell-line adapted isolates



`closed' configuration

`open' configuration

1993; Groenink *et al*, 1992). This may be mediated through exposing the V3 loop and CD4 binding-site.

7.7.3 The Implications of the dual configuration hypothesis upon neutralisation studies

The dual configuration hypothesis has a number of important implications: firstly, that the V3 loop may not be the principal neutralising determinant (PND) of the viruses found *in vivo*. Indeed, the V3 loop was defined as the PND following the experimental inoculation of animals with synthetic peptides representing the V3 loop (Ruche *et al*, 1988). Many of the neutralising assays used the 'neutralisation-sensitive' isolates such as HIV-1_{IIIB}. It may be that other neutralising epitopes are more important targets of neutralisation *in vivo*. In one study, a primary isolate was found to be more sensitive than a laboratory-adapted strain to neutralising antibodies directed to the V2 region (Gorny *et al*, 1994).

Secondly, the hypothesis suggests that studies on the immune response to HIV infection should be conducted using primary isolates. The properties of primary isolates have been examined in a number of recently published work (Iversen *et al*, 1995; Gao *et al*, 1996; Sullivan *et al*, 1995) and has formed the basis of the work described in this thesis.

Thirdly, the evolution of the V3 loop of HIV-1 *in vivo* may be restricted by the presence of a strong immune system which selects for a V3 loop which is less accessible to neutralising antibodies, ie. has a 'closed' configuration. Recent studies have shown that the V3 loop sequences early in infection, and those in non-lymphoid tissues, show a restricted sequence variation, a low net charge, and are predicted to confer a macrophage, NSI phenotype (Zhang *et al*, 1993; Zhu *et al*, 1993; Donaldson *et al*, 1994b; Korber *et al*, 1994; Milich *et al*, 1993). However, a number of researchers have shown that highly divergent V3 loops may be found within the plasma of a single individual, during the asymptomatic phase (Leigh Brown and Clelland, 1996).

Moore and Ho, 1995, have suggested that syncytium-inducing isolates only appear later in infection, because a strong neutralising antibody response to these variants selects against these variants during the early stages of infection. During the latter stages of HIV-1-infection, a compromised immune system with a reduction in titres of neutralising antibodies, will allow viruses with a higher replication rate but increased sensitivity to neutralising antibodies to emerge. With respect to this, Bonhoeffer *et al*, 1995, have found that the V3 sequences with an inferred SI phenotype from patient 82, have a significantly lower K_s/K_a (ie. that selection is stronger) than sequences of an inferred NSI phenotype. This supports the hypothesis that the immune system has a greater effect against SI-type viruses than NSI/macrophage-tropic viruses.

7.7.4 Approaches to Examining Autologous Neutralising Antibody Responses

There are two main experimental approaches taken to study autologous neutralising antibody responses to the evolving virus population found *in vivo*.

The first, is to obtain viral isolates from sequential peripheral blood samples, and to examine the neutralising properties of autologous serum samples taken at time points before and after the isolates were obtained. Such an approach has been taken by Albert *et al*, 1990; Arendrup *et al*, 1992; Tremblay *et al*, 1990; Moore *et al*, 1994b, Cao *et al*, 1995) These studies have shown that isolate-specific neutralising antibodies to autologous viral strains arise after seroconversion, and that in some cases, subsequent isolates are resistant to neutralisation from early sera. In one study, neutralising antibodies to the resistant forms developed, however, the titre of these antibodies was lower than to the earlier isolates (Arendrup *et al*, 1992)

However, most viral isolates are genetically heterogeneous (Hahn *et al*, 1986; Meyerhans *et al*, 1989; Delwart *et al*, 1993), as such, the neutralising antibody titres obtained will be to a mixture of variants which may have different neutralising sensitivities, and not to a specific variant. In these cases it will not be possible to correlate resistance with a specific *env* sequence. Another source of difficulty is

obtaining isolates in the early stages of infection when viral load is very low and isolation difficult.

A second approach is to engineer recombinant viruses expressing the *env* sequences whose neutralising sensitivities are to be examined. This approach has been taken to examine the autologous neutralising responses following the accidental infection of a laboratory worker with HIV-1_{IIIB} (Di Marzo Varonese *et al*, 1993), and following the experimental infection of monkeys with HIV-1 and SIVmac (Burns *et al*, 1993; Back *et al*, 1993; Reitz *et al*, 1994).

Di Marzo Varonese *et al*, 1993, examined the neutralising response following the accidental infection of a laboratory worker with HIV-1_{IIIB}. A predominant *env* variant, detected one year after infection, was introduced into the pHXB2 clone (an infectious clone derived from the HIV-1_{IIIB} isolate, Shaw *et al*, 1984, and used in this study), and this clone was found to be resistant to a panel of monoclonal antibodies directed to the V3 loop, and to neutralisation by autologous sera. However, this variant was detected *after* the development of a neutralising response capable of neutralising the variant and persisted in all subsequent isolates (Reitz *et al*, 1994). This suggests that a variant specific neutralising antibody response did not select this variant. The predominance of this resistant variant may have occurred by chance, or by antibody-enhancement of infection (Sanchez-Palomino *et al*, 1993; Domingo *et al*, 1993). Also, infection was with the laboratory strain HIV-1_{IIIB}, which may not be representative of the immune response to infection with primary isolates, which as described above, are less sensitive to antibody neutralisation.

The recombinant approach was used by Burns *et al*, 1993, who studied the sequence-specific immune response in rhesus monkeys infected with the pathogenic clone SIV_{mac239}. Two variant *env* sequences, arising at 69 and 93 weeks post-infection, respectively, were substituted back intothe parental clone SIV_{mac239} to generate two variant viral clones. Sequential sera from the monkey infected with SIV_{mac239} showed much stronger neutralising titres to the parental virus, than to either of the two variants. Experimental infection of naive monkeys with the two variants generated high titres of type-specific neutralising antibodies, showing that the two variants were not inherently

less immunogenic. The authors concluded that SIV_{mac239} mutants resistant to serum neutralisation arise during the course of infection in rhesus monkeys.

Similarly, following the experimental infection of chimpanzees with $HIV\text{-}1_{IIIB}$, recombinant viruses chimeric for the *env* gene, were generated to examine the sequence-specific neutralising antibody responses (Nara *et al*, 1990; Back *et al*, 1993). These studies have shown that viruses resistant to V3 antibodies arise during infection in response to a V3-specific neutralising antibody response. In one instance, the mutation conferring resistance mapped to gp41 (Back *et al*, 1993). However, these studies were of the experimental infection of chimpanzees with the laboratory-adapted isolate $HIV\text{-}1_{IIIB}$. Again, such a result with the laboratory-adapted isolate may not be representative of the immune response to primary isolates.

In more recent work by Sullivan *et al*, 1995, recombinant viruses were generated to examine the structure and neutralising sensitivity of different *env* genes: two from laboratory isolates, and three from primary isolates. These studies did not address the question of the development of 'escape' mutants to neutralising antibodies, but confirmed the greater sensitivity of laboratory-adapted isolates to neutralisation.

7.8 Summary of the Results of this Study

An HIV-1 cassette-vector has been made to allow examination of the functional properties of variants of the *env* gene. The vector, pHXB2-MCS, is modified from the infectious proviral clone pHXB2-D (Fisher *et al*, 1985), and has unique cloning sites for exchange of the *env* gene.

Initially, ten recombinant clones were constructed by replacing a 1.7kb fragment of pHXB2-MCS with ten variants of the *env* gene. The *env* sequences were cloned directly from the peripheral blood samples of a single infected individual. The sequences are from four time points and are representative of the spectrum of variants within this individual.

On transfection into T-cell lines and COS-PBMCs, virions recovered from two clones (G.1, H.1) were able to replicate. Each transfected culture showed a burst of p24 72hr post-transfection, a transient decrease and subsequent increase in p24 over a period of 30 days. Each induced the formation of syncytia and could be passaged in a variaety of T cell lines and in PBMCs. Virions recovered from another two clones (E.1, F.1) showed evidence of low level replication. An increase in p24 could be detected in the first week of culturing, but this level declined to <1ng/ml by the second week. Virions from clones E.1 and F.1 could not be passaged in PBMCs or a Variety of T-cell lines.

In contrast, virus replication could not be detected from the remaining six clones (A.1, B.1, C.1, D.1, I.1, J.1). After an initial burst of p24 72hr post-transfection, levels dropped to below <1ng/ml within a week. Virus replication could not be detected following passage in either T-cell lines or PBMCs.

The remaining work of this study has been directed towards understanding why such a small proportion of clones are replication competent.

7.9 Discussion of Results

Perhaps the most obvious explanation for the low frequency of replication-competent clones is that the non-productive clones contain inactivating mutations. However, extensive sequencing has been carried out for each clone, including across each of the five hypervariable regions and across part of gp41 (chapter 5; section 5.4). This totalled 16kb of sequencing but did not reveal any inactivating mutations. It is, however, conceivable that mutations lie within the unsequenced regions of the *env* gene. But if there is a high incidence of inactivating mutations, we would expect these to show a random distribution across the *env* gene and not be limited to the unsequenced regions.

To investigate further the presence of inactivating mutations, the full sequence of two of the non-infectious clones (A.1 and C.1) was obtained (chapter 5; section 5.4). Each was found to encode a full open reading-frame with no obvious inactivating mutations. A more detailed examination of the possibility of inactivating mutations in these clones,

is described separately below (section 7.10).

Closer inspection of the deduced amino acid sequences of the third hypervariable region (V3) from each of the ten clones, reveals an interesting result: the two fully-infectious clones have V3 loops predicted to confer an SI and a T-cell-line tropic phenotype. In contrast, the non-infectious clones and the two partially infectious clones have V3 loop sequences predicted to confer an NSI and a non-T-cell-line phenotype (chapter 5; section 5.5). Firstly, this result goes against the hypothesis that the majority of the non-infectious clones contain inactivating mutations - if there are inactivating mutations in these clones, we would not expect these mutations to be limited to clones of an NSI-type. In addition, the non-infectious clones are derived from the early (1987) blood sample and the later sample (1990). The infectious clones are both from 1989 samples. Again, it would be difficult to see why inactivating mutations should limited to clones these two sample dates.

Isolates that are 'NSI' have a number of biological properties that are different from 'SI' isolates (Asjo *et al*, 1986; Fenyo *et al*, 1989). In general, NSI isolates can grow in monocyte-derived macrophage cultures, T-cells, but not T-cell-lines. They frequently have a lower replicative ability, that is, produce p24, or reverse transcriptase (RT), at a lower rate *in vitro* than SI isolates. In contrast, SI isolates generally grow poorly in monocyte-derived macrophages, but replicate well in T-cells and particularly well in T-cell-lines.

It is conceivable, then, that a similar difference in replicative capacity exists between the 'NSI-like' clones and the SI clones of this study, which renders the NSI-like clones very slow growing *in vitro*. Slow/low viruses that grow in T-cells but cannot be passaged, have been described by Asjo *et al*, 1986 and Fenyo *et al*, 1989. These viruses may be similar to the three poorly-growing clones of this study (E.1, F.1 and P.2). However, the remaining NSI-like viruses of this study, did not show convincing evidence of virus replication.

It may be that the vector, pHXB2-MCS, is unable to support the growth of NSI-like sequences. One way to examine whether pHXB2-MCS is able to produce replication-

competent viruses with NSI sequences, would be to substitute the *env* gene from a well-characterised NSI clone such as YU2, ADA, SF162 or 89.6 (Li *et al*, 1992; Cheng-Mayer *et al*, 1991, Collman *et al*, 1992) into pHXB2-MCS. We were, however, unable to do this, partly due to the unavailability of these clones, but also due to technical difficulties with cloning into pHXB2-MCS. In addition, many well-characterised NSI isolates are studied because of their high replicative ability, and may not represent NSI-like primary isolates.

If such an incompatibility exists between NSI-like sequences and pHXB2-MCS, we hypothesised that it may be due to the formation of a hybrid envelope protein arising from the substitution of the 1.7kb *env* fragment (cassette 1). In the ten '1.7kb' clones, the gp41 transmembrane protein is hybrid between the pHXB2-MCS and the patient sequences (chapter 3: figure 3.2); residues from the HXB2 gp41 part may be unable to interact correctly with the patient gp41 and gp120 sequences.

To prevent a hybrid gp41, two further clones with patient-derived NSI-like sequences were constructed. The first (P.2), has the whole of the *env* gene derived from the patient, and the second (A.3), has just gp120 derived from the patient (chapter 4; sections 4.2.4. and 4.2.5). Each has a V3 loop predicted to confer an NSI phenotype. The latter clone (A.3) was found to be non-infectious, but clone P.2 was minimally-infectious, but could not be passaged.

To test further the hypothesis that SI-like sequences can form infectious viruses with pHXB2-MCS, four more clones, this time with SI-like sequences, were constructed (K.1, L.1, M.1, N.1; chapter 4, section 4.2.3). These were constructed with *Taq* amplified *env* sequences from the 1989A DNA sample. However, virus replication could not be detected following transfection into COS-PBMCs or T-cell lines. Although these results do not support the 'SI/NSI' hypothesis, these clones may have incorporated inactivating mutations due to the high rate of mutations associated with *Taq* polymerase.

Three clones (Q.2, R.2, S.2) were generated by exchanging the whole of the *env* gene for sequences obtained from the 1989A DNA sample. Each of these produced infectious viruses. The V3 loop of one (R.2) has been obtained and shows an SI-like sequence

(chapter 5; section 5.5) supporting the hypothesis that only SI sequences are replication competent.

However, in a final set of experiments we were unable to detect infectious viruses from two clones incorporating a 1.7kb fragment of the *env* gene of the infectious, SI, molecular clone $SF2_{MC}$. Similarly, the 2.5kb provial fragment from $SF2_{MC}$ expressed on the pHXB2-MCS background failed to support virus replication, even though this fragment contains the entire *env* gene from $SF2_{MC}$. An inactivating mutation may have been introduced during the cloning, or there may be a requirement for accessory genes absent in pHXB2-MCS.

Indeed, with each of the non-infectious clones, a deficiency may reside with pHXB2-MCS, which has inactivating mutations in *vpu*, *vpr* and *nef* genes (Myers *et al*, 1991). These 'accessory' genes may be required to form infectious particles with the NSI-like *env* sequences of this study and with $SF2_{MC}$ *env* gene. The role of accessory genes is discussed further in section 7.11, below. The most direct way to test this hypothesis would be to express the *env* genes on another full-length HIV-1 clone, such as HIV-1_{YU2}, HIV-1_{SF162}, HIV-$1_{86.9}$, HIV-$1_{JR\text{-}CSF}$ (Li *et al*, 1992; Cheng-Mayer *et al*, 1990; Collman *et al*, 1992). Again this was thwarted due to cloning difficulties, time restraints and poor accessibility of published molecular clones.

7.10 Discussion on the Functional Ability of *env* Genes Amplified Directly from PBMC DNA

Although no inactivating mutations were found by sequence analysis, if the non-productive clones contain inactivating mutations then it will be important to recognise where these came from. There are two main sources of inactivating mutations: firstly, mutations could have been introduced into the *env* genes during the process of PCR amplification and cloning (ie. **artefactual error**), or secondly, the *env* variants as they were present in the PBMCs prior to amplification, could be defective (ie. **defective proviruses**).

7.10.1 Artefactual Error

One of the most obvious sources of artefactual error is the error-prone process of PCR-amplification used to amplify the *env* genes from PBMC DNA. Errors may have arisen by nucleotide mis-incorporation, or by the production of artefactual hybrid molecules by recombination *in vitro* (Meyerhans *et al*, 1990). In this study, the mis-incorporation of nucleotides was reduced through the use of *Pfu* DNA polymerase in place of the more commonly used *Taq* DNA polymerase. The error rate of *Pfu* polymerase, calculated in this study, was less than one error in 16kb of amplified and cloned DNA. This error-rate was considered low enough to allow the amplification of *env* gene sequences from PBMC DNA without the introduction of errors. The production of artefactual recombinant molecules was reduced by amplifying each proviral sequence in a separate reaction tube (chapter 4).

Artefactual errors could also occur through the two cloning steps subsequent to the PCR amplification, that is, the cloning of the amplified *env* gene sequences into the subcloning vector (pNBXXΔenv), and the transfer of the cloned env genes to the proviral vector (pHXB2-MCS). There are two main sources of error associated with cloning: firstly, the exposure of the DNA to mutagenic UV light - required in this study to isolate the cloned *env* gene sequences prior to cloning into the proviral vector; and secondly, the propagation of the cloned sequences in *E.coli*, where it has been shown that HIV-1 proviral sequences can often be unstable and may contain bacterial insertion sequences (Kusumi *et al*, 1992).

Point mutations introduced during the cloning steps may have gone unnoticed, although large DNA re-arrangements would have been detected following the restriction-mapping of each clone. In this study, the *E.coli* strain DL655F' (chapter 2, section 2.4) was found to be capable of stably propagating both *env* gene subclones, and the larger full-length proviral clones.

7.10.2 Defective Proviruses

It is also possible that the cloned *env* gene variants were derived from a subset of defective proviral sequences that were present in the original PBMC sample. It has often been argued from the frequent detection of stop codons from clones amplified from PBMC DNA, that many of the proviral sequences found *in vivo* are defective (Meyerhans *et al*, 1989; Kusumi *et al*, 1992; Li *et al*, 1992; Sabino *et al*, 1994). However, most of these reports used *Taq* polymerase to amplify the sequences, which as has been described above, is associated with a high error rate. In other studies, very few inactivating mutations were found in the PBMC population (Balfe *et al*, 1990, Burns and Desrosiers, 1991).

It has also been argued from the observation of the long term survival of HIV-1 variants in the PBMC population, that these may be defective proviruses (Simmonds *et al*, 1991; Haigwood *et al*, 1993; Zhang *et al*, 1993; Kusumi *et al*, 1992). For example, in the analysis of Simmonds *et al*, 1991, of the V4 and V5 hypervariable regions of the *env* gene from sequential PBMC and plasma samples, it was found that proviral variants may be found for up to three years after they were found in the plasma (see chapter 5). From this observation, the authors have argued that many PBMCs represent long-term infected cells, which survive infection because they harbour defective proviruses. Indeed, as described in chapter 5, many of the cloned variants obtained in this study were cloned from PBMC DNA samples at time-points when the corresponding variants were not found in the plasma. It could be argued, therefore that these clones were derived from defective proviral sequences. However, three of the viable clones constructed in this study: 82.24 (G.1), 123.14 (H.1) and 82.04 (R.2), were cloned from PBMC samples taken at dates in which the equivalent sequence was not found in the plasma. Therefore, at least in these cases, this argument does not hold.

There have been a number of investigations determining the proportion of infectious proviruses in PBMCs (Brinchman *et al*, 1991; Simmonds *et al*, 1990a; Bienasz *et al*, 1993). These studies were carried out by measuring the infectious viral load in PBMCs by limiting-dilution culture, and comparing this figure to the number of proviral DNA

copies in the original sample, calculated by a quantitative PCR method. By assuming that each infected PBMC harbours a single provirus, (the figure estimated by Simmonds *et al*, 1990a and Brinchman *et al*, 1991), it is possible to produce a ratio of the number of infectious proviruses to the number of proviral sequences. In the study of Brinchman *et al*, 1991 and Simmonds *et al*, 1990a, the ratio was found to be high - ranging from between one infectious provirus to every five proviral DNA copies (1:5), to one infectious copy for each proviral sequence (1:1). From these studies it would appear that a high proportion of proviruses found in PBMCs are infectious. However, in a more extensive study of Bienasz *et al*, 1993, a much wider variation was found, of between one infectious provirus to each proviral copy (1:1), to one infectious copy for every 1000 proviral copies (1:1000). Clearly, therefore, it is possible that the large majority of proviruses are incapable of producing infectious viruses. However, the number of infectious proviruses calculated by limiting-dilution culture, may be an underestimate. It may be the case that not all viable viruses can replicate to detectable levels in culture. In addition, the presence of CD8+ T lymphocytes in the PBMC cultures, has been shown to suppress the expression of HIV-1 (Cocchi *et al*, 1995). This would lead to an underestimate of the number of infectious proviruses present.

7.10.3 Alternative Approaches to Obtaining *Env* Genes from Patient Material

If the majority of PBMC-derived proviral genomes are defective, then it may be necessary, in future work, to screen the cloned *env* genes for an intact-reading frame, or indeed, clone the *env* genes from another source.

It may be possible to screen for an intact reading-frame using *in vitro* transcription/translation systems (as described in chapter 4). Alternatively, eukaryotic expression vectors such as pSVIII*env*CAT (Helseth *et al*, 1990a), or a vaccinia virus expression vectors could be used to screen for a full-length gp160.

PBMCs may be cultured for a limited number of passages, in order to increase the

proportion of infectious proviruses prior to cloning (Groenink *et al*, 1991). Culturing should stimulate virus replication and increase the proportion of viable viruses. However, a major disadvantage of this approach is that culturing greatly reduces the variation of the virus population (Kusumi *et al*, 1992; Sabino *et al*, 1994). Therefore, following culturing, it is unlikely that the full-spectrum of variants will be represented.

An alternative approach would be to carry out limiting-dilution culturing, as has been described by Schuitemaker *et al*, 1992; Connor *et al*, 1993a and Bienasz *et al*, 1993. Through limiting-dilution, each HIV-1 positive culture should be initiated from a single infectious unit, and this should reduce competition between variants and the selection of the fastest-replicating viruses. However, this method selects those variants best adapted to culture conditions. Very low replicating variants may be missed.

Another source of *env* genes is the virion-associated RNA found in the plasma. It has been argued that plasma variants will represent actively replicating sequences, and therefore, should contain a higher proportion of infectious viruses (Simmonds *et al*, 1991). However, it is currently, exceedingly difficult to amplify and clone long stretches of genomes from an RNA source.

7.11 Discussion on the use of pHXB2-MCS as a Cassette-Vector

If the sequences of *env* genes must be SI-like to produce rapidly-growing viruses in the recombinant system developed in this study, then this may be due to the choice of vector (pHXB2-D). pHXB2 is a laboratory-adapted, SI-virus (Fisher *et al*, 1985), so it may be expected to support only the growth of SI-variants.

It is conceivable that certain *env* sequences, and in particular NSI-types, may not interact with the genetic elements provided by pHXB2-MCS required to form infectious virions.

One of the most obvious problems of pHXB2 as an expression vector, is that it has

inactivating mutations in three accessory genes: *vpu*, *nef* and *vpr*. The Vpr protein is truncated - 25 amino acids shorter than other subtype B isolates due to a frameshift mutation. The *vpu* gene has no start codon, and the *nef* gene contains a premature termination codon (Myers *et al*, 1991). As has been demonstrated in this study, these three accessory genes are dispensable for the growth of HIV-1 viruses in cultures of PBMCs, CD4-positive T-cell-lines, and in CD4-positive T-lymphocytes (Fisher *et al*, 1985; Westervelt *et al*, 1992b). However, recombinant studies have shown that the *vpu* and *vpr* genes are essential for growth in monocyte-derived macrophage cultures (Gibbs *et al*, 1994; Westervelt *et al*, 1992b). These two genes have a complementary function in macrophage culture, such that a functional copy of only either of these genes is sufficient for productive infection. Vpr contains a nuclear localisation signal, which connects the pre-integration complex to the cell nuclear import pathway, which is necessary for replication in non-dividing cells, such as the macrophage (Heinzinger *et al*, 1994).

It could be hypothesised that the lack of *vpu* and *vpr* in pHXB2-MCS, will prevent productive infection with certain *env* variants, if these variants are only capable of productive infection in monocyte-derived macrophages. However, the majority of primary isolates are 'dual tropic' and can productively infect both macrophage cultures *and* PBMCs and CD4+ T-lymphocytes (Connor *et al*, 1993b, 1994; Schuitemaker *et al*, 1991, 1992 and Valentin *et al*, 1994). Therefore, it is unlikely that the absence of *vpu* and *vpr* will have prevented the replication of the viruses in this study which have predicted NSI/macrophage-tropic phenotype, because these should grow in PBMCs, and as such, should not show a requirement for *vpu* and *vpr*.

The proviral clone pHXB2 has been used quite extensively by other groups to generate recombinant clones to examine the function of the accessory genes (Westervelt *et al*, 1992b; Fisher *et al*, 1986), and properties of various HIV-1 *env* genes (Sullivan *et al*, 1995; De Jong *et al*, 1992a and 1992b). Several of these have shown that pHXB2 can generate infectious viruses with full-length *env* gene sequences of an NSI/macrophage (non-T-cell-line) tropic phenotype, without altering the defective genes *vpr*, *vpu* and *nef* (Westervelt *et al*, 1992b; Sullivan *et al*, 1995). For example, the *env* gene from the prototype macrophage-tropic clone 'ADA', when expressed on an HXB2-background,

forms recombinant viruses that grow well in PBMCs but not macrophages (Westervelt *et al*, 1992b). Sullivan *et al*, 1995, used a defective virus pHXB*env*CAT - an *env* gene-defective derivative of pHXB2 (Helseth *et al*, 1990a), to assess the ability of a number of envelope glycoproteins (from both primary and laboratory-adapted strains), to complement virus entry. When the *env* genes were transfected together with pHXB*env*CAT, recombinant viruses capable of infecting CD4+ T-lymphocytes, and SupT1 cells, were generated. Five envelope glycoproteins were used: two from primary macrophage-derived/NSI clones 'ADA', 'YU2' (Li *et al*, 1992), and two from laboratory-adapted isolates HXB2, 'MN' (Shaw *et al*, 1984; Fisher *et al*, 1985), and one from an unusual cytopathic macrophage-tropic isolate '89.6' (Collman *et al*, 1992). Although this study used the *env* gene of primary isolates, each clone was selected for its high replicative ability to facilitate the production of virus stocks, and may not be representative of the majority of viruses found *in vivo*.

Iversen *et al*, 1995, exchanged the entire *env* gene of molecular clone pNL4-3 (which is also from the same isolate as pHXB2-D: HIV-1_{IIIB}) with four sequences amplified directly from PBMC DNA, and this resulted in the production of four non-infectious viruses. Replication could not be detected in peripheral blood lymphocytes nor monocyte-derived macrophages. Each clone had an open reading frame in gp160, but an attenuating point mutation in the *rev* gene contained within the exchanged region.

In this study, the cassette-vector pHXB2-MCS was designed specifically to allow exchange a 1.7kb fragment of the *env* gene. This fragment does not include any open reading frames of *rev*, *tat*, *vpu* or *nef* (chapter 3).

In our pHXB2-MCS system, PBMC-derived SI-sequences produced rapidly-growing viruses when the whole *env* gene was replaced (cassette-2) and when the *env* gene was hybrid (cassette-2 - with gp120 and the N-terminal third of gp41 exchanged). The non-infectious or slowly-growing viruses were constructed with either a hybrid *env* gene (cassette-1 and -3) or a full *env* gene (cassette-2) from an NSI variant (chapter 6). In this case, it would appear that the region exchanged (cassette-1, -2 or -3) is not a major factor contributing to the infectivity of the resultant recombinant viruses.

It could also be argued that the recombinant viruses expressing predicted-NSI sequences are not capable of growth in the cell types in which we have attempted to propagate them, namely PBMCs and transformed CD4+ cell lines.

We have no information concerning the cell type in the PBMCs from which the sequences were derived. The DNA used to obtain the sequences was extracted from unfractionated buffy-coat. The cell types infected *in vivo* in peripheral blood appear to be mainly CD4+ T lymphocytes of the memory cell subtype (Schnittman *et al*, 1989 and 1990). Other reports have indicated the presence of HIV-1 sequences in monocytes and dendritic cells but only at very low levels (Knight *et al*, 1990). It is conceivable therefore, that the *env* sequences obtained in this study, were derived from viruses showing exclusive tropism for monocytes, or for dendritic cells. These cells may form such a low proportion of PBMCs that viral replication cannot occur to detectable levels.

With the recent discovery of the co-receptors required for HIV-1 infection, CD4+ cell lines expressing these receptors may facilitate virus infection and replication.

Whilst it is clear from the work described in this thesis, that the construction of viable recombinant viruses with *env* genes taken directly from PBMC is difficult, this type of approach will be necessary in studying the property of viruses, especially where the isolation through cell culturing is not possible or is undesirable.

For example, to study virus found in different organs at post-mortem from which virus isolation is difficult. It is, similarly, difficult to obtain viral isolates from patients with low viral load, such as is frequently found in patients at the early asymptomatic stage of infection, or in long-term non-progressors where proviral load is often measured as below 1 copy per 20,000 PBMCs (Haigwood *et al*, 1993; Cao *et al*, 1995). These viruses may show interesting *in vitro* properties, which confer a low level replication *in vivo* and an apparent slow CD4 decline.

Appendices

Appendix A Solutions and Media

2xTY-broth
Per liter:
16g tryptone (Oxoid)
10g yeast extract (Oxoid)
5g NaCl (Sigma)
Adjust to pH 7.0 with 10M NaOH

For 2xTY-agarose plates:
10g agarose (BRL)

TE
10mM Tris, pH 8.0
1mM EDTA, pH 8.0

TBE 10x
89mM Tris-base
89 mM boric acid
3mM EDTA

TfbI
30m mM potassium acetate
100 mM RbCl
10 mM $CaCl_2$
50 mM $MnCl_2$
15 % glycerol
adjust to pH 5.8 with 0.2 M acetic acid and sterilise by filtration.

TfbII
10 mM PIPES
10 mM RbCl
75 mM $CaCl_2$
15% glycerol
Adjust to pH 6.5 woth KOH and sterilise by filtration.

STET
8% w/v sucrose
0.5% v/v Triton X-100 (Sigma)
50 mM EDTA pH8.0
50 mM Tris pH 8.0

Ficoll Stop Buffer (FSB)
100 mM EDTA
20% w/v Ficoll Type 400
0.05% w/v bromophenol blue

6% Polyacrylamide Sequencing gel

Ligase 10x buffer (Boehringer)
500mM Tris, pH 7.6
100 mM $MgCl_2$
10 mM ATP

KGB (K^+ Glutamate Buffer) 10x (McClelland *et al*, 1988)
1M potassium glutamate
100mM Mg acetate
500ug/ml bovine serum albumen
5mM -mercaptoethanol

filter sterilise and store at 4^0C

Appendix B Restriction Sites of HXB2R

HXB2R RESTRICTION SITES

SEQUENCE RANGE: 1 to 9265

Enzyme	#Cuts	Positions							
Aat1	2	4953	6380						
Acc1	8	506	1225	2474	3936	4377	4624	4879	5333
Acc3	1	8939							
Afl2	4	63	4289	5180	9148				
Afl3	3	5601	6051	7044					
Aha1	3	1402	6428	8868					
Aha3	11	790 6156	886 6795	1336 8614	2088	3133	3262	4080	4330
AlwN1	6	693	4938	7462	7546	8237	9052		
Apal	1	1556							
Apal1	1	6160							
Asel	2	5043	6651						
Asp718	4	3372	3700	5893	8561				
Aval	3	7939	8443	8926					
Ava3	6	797	2437	3601	3927	6084	6292		
Avr2	3	1557	4977	5207					
Ball	2	2167	4099						
BamH1	1	8021							
Ban3	1	377							
Bbel	1	187							
Bell	1	1975							
Bg12	6	19	1642	6587	7167	8597	9104		
Bsml	3	1177	7606						
BspMl	2	3864	4592						
BspM2	1	8939							
BssH2	1	257							
BatXI	5	1181	1920	2550	2778	3954			
Bau361	4	1800	5502	6862	8557				
Corl	1	8443							
Cfrl	3	1403	2165	4097					

Cfr10I	2	1228	4139						
ClaI	1	377							
CvnI	4	1800	5502	6862	8557				
DraI	11	790 6156	886 6795	1336 8614	2088	3133	3262	4080	4330
Dra3	2	6146	8843						
EcoN1	2	5517	7182						
EcoR1	2	4194	5289						
EcoR5	3	2525	8666	8745					
EcoT221	6	797	2437	3601	3927	6084	6292		
EspI	1	8410							
HincII	4	3044	3067	6334	6348				
HindIII	6	77	631	1258	5573	7687	9162		
HphI	10	301 7003	769 7899	2441	2553	3487	4129	4520	4616
KpnI	4	3376	3704	5897	8565				
MroI	1	8939							
MstII	4	1800	5503	6852	8557				
NarI	1	184							
NcoI	1	5221							
NdeI	2	4669	5950						
NheI	1	6806							
NsiI	6	797	2437	3601	3927	6084	6292		
PaeR7I	1	8443							
PfiMI	2	3038	4849						
PflMI	2	3038	4849						
PstI	1	965							
PvuII	6	693	2847	3891	6631	8592	9066		
SacI	4	37	228	5550	9122				
SalI	1	5332							
Sau3AI	4	1800	5502	6862	8557				
ScaI	7	2299	2418	3771	6949	6964	6973	8946	
SpeI	1	1053							
SphI	1	993							
SspI	4	2571	5705	7114	8153				
SstI	4	37	228	5550	9122				
StuI	2	4953	6380						

Xho1	1	8443			
Xmn1	4	388	822	4038	4739

Non Cutters

Aat2

Aos1

Asu2

Bg11

BspH1

BstB1

BstE2

Eag1

Eco47III

Eco521

Fdi2

Fsp1

Hpa1

Mlu1

Mst1

Nae1

Not1

Nru1

Pvu1

Sac2

Sma1

SnaB1

Spl1

Sst2

Tth111I

Xba1

Xma1

Xma3

Appendix C Fragment sizes produced upon single-enzyme digestion of full-length recombinant proviral clones

Key of restriction patterns following sinlge digest with *Bsm*I, *Hin*dIII, *Kpn*I, *Ssp*I

clone	BsmI	HindIII	KpnI	SspI
pHXB2.MCS	H	H	H	H
A	2A	H	H	H
B	2A	H	H	H
C	H	H	H	H
D	2A	1A	1A	H
E	2A	H	H	H
F	2A	1A	H	H
G	1A	1A	H	H
H	1A	1A	H	H
I	H	H	1B	1A
V	H	1A	H	1B
J	1B	H	H	H
P	2A	1A	H	2A
Q	2A	1A	H	2A
R	1A	1A	1A	2A
S	2A	1A	H	2A

H, 1A, 1B and 2A are (arbitrary) names for restriction-fragment patterns. (Patterns labelled H are identical to pHXB2-MCS. 1A and 1B are patterns produced when there is a single site within the cassette-region, and patterns labelled 2A are those generated when their are two sites within the cassette).

BsmI

H	1A	1B	2A
6429 (V) ~6100 (C) 2500 (V)	~6100 (C) 5004 (V) 3952 (V)	8929 (V) ~6100 (C)	~6100 (C) 5004 (V) 2500 (V) 1410 (V)

HindIII

H	1A
4315 (C)/~4300 (C) doublet ~2500 (C) 2114 (V) ~1800 (C) 1475 (C) 627 (V) ~550 (C)	4315 (C)/~4300 (C) doublet ~2500 (C) ~1800 (C) 1599 (V) 1475 (C) 627 (V) ~550 (C) 531 (V)

KpnI

H	1A	1B
~10050 (C) 2668 (V) 2193 (V) 328 (C)	~10050 (C) 4885 (V) 328 (C)	~10050 (C) 3161 (V) 1641 (V) 328 (C)

SspI

H	1A	1B	2A
~3300 (C) 3134 (C) ~2500 (C) ~1800 (C) 1409 (V) 1039 (V) ~550 (C)	~3300 (C) 3134 (C) ~2500 (C) ~1800 (C) /1870 (V) doublet ~600 (V) ~550 (C)	~3300 (C) 3134 (C) ~2500 (C) 2448 (V) ~1800 (C) ~550 (C)	~3300 (C) 3134 (C) ~2500 (C) ~1800 (C) 1409 (V) 650 (V) ~550 (C) ~350 (V)

(V) Vector: are restriction-fragments mapping outside the cassette, and are therefore constant.

(C) Cassette: are the variable, restriction-fragments contained entirely within, or spanning, the cassette.

Appendix D Nucleotide Sequences of the *env* Gene Clones

Taq-amplified sequences:

a) V4-C4 (primer 407 sense)

```
clone          1                                                      50
P0821013407    CTTGGAATTC AACATGGGAT TTAACACAAC TTAATAGTAC TCAGAATAAA
P0821022407    CTTGGAATTC AACATGGGAT TTAACACAAC TTAATAGTAC TCAGAATAAA
P0820611407    .......... .....gggat ttaacacaac ttaatagtac ccagaataaa
P0820612407    .......... .....GGGAT TTAACACAAC TTAATAGTAC CCAGAATAAA

               51         *                                    *     100
P0821013407    GAAGAAAATA TCACACTCCC ATGTAGAATA AAACAAATTA TAAACATGTG
P0821022407    GAAGAAAATA ACACACTCCC ATGTAGAATA AAACAAATTA TAAACATGTG
P0820611407    gaagaaaata tcacactccc atgtagaata aaacaaatta taaacatgtg
P0820612407    GAAGAAAATA TCACACTCCC ATGTAGAATA AAACAAATTA TTAACATGTG

               101                                                    150
P0821013407    GCAGGAAGTA GGAAAAGCAA TGTATGCCCC TCCCATCAGA GGACAAATTA
P0821022407    GCAGGAAGTA GGAAAAGCAA TGTATGCCCC TCCCATCAGA GGACAAATTA
P0820611407    gcaggaagta ggaaaagcaa tgtatgcccc tcccatcaga ggataaatta
P0820612407    GCAGGAAGTA GGAAAAGCAA TGTATGCCCC TCCCATCAGA GGACAAATTA

               151                                                    200
P0821013407    GATGTTCATC AAATATTACA GGGCTACTAT TAACAAGAGA TGGTGGTAAT
P0821022407    GATGTTCATC AAATATTACA GGGCTACTAT TAACAAGAGA TGGTGGTAAT
P0820611407    gatgttcatc aaatattaca gggctactat taacaagaga tggtggtaat
P0820612407    GATGTTCATC AAATATTACG GGGCTACTAC TAACAAGAGA TGGTGGTAAT

               201                                                    250
P0821013407    AGTGGTAACA AAAGCAACG. .......... .......... ..........
P0821022407    AGTGGTAACA AAAGCAACGA CACCACCGAG ACCTTCAGAC CTGGGGGAGG
P0820611407    agtggtaaca aa........ .......... .......... ..........
P0820612407    AGTGGTAACA AAAGCAA... .......... .....-.... ..........

               251                      276
P0821013407    .......... .......... ......
P0821022407    AGATATGAGG GACAATTGGA GAAGAG
P0820611407    .......... .......... ......
P0820612407    .......... .......... ......
```

b) V3 region (primer 307 antisense)

```
clone          1                                                  *   50
P0820612307    TCTTAATTTT ATAACTATCT GTTTTAAAGT TTCATTCCAT TTTGCTCTAC
P0820611307    TCTTAATTTT ATAACTATCT GTTTTAAAGT TTCATTCCAT ATTGCTCTAC

               51                                                 *   100
P0820612307    TAATGTTACA ATGTGCTTGT CTTATATTTC CTATTATTTG TTCTGTTGTA
P0820611307    TAATGTTACA ATGTGCTTGT CTTATATTTC CTATTATTCG TTCTGTTGTA

               101                                                    150
P0820612307    TAAACTGCTC TCCCTGGTCC TATATATATC CTTTTTCTTG TATTGTTGTT
P0820611307    TAAACTGCTC TCCCTGGTCC TATATATATC CTTTTTCTTG TATTGTTGTT

               151            168
P0820612307    GGGTCTTGTA CAAGTAAT
P0820611307    GGGTCTTGTA CAAG....
```

Pfu-amplified sequences:

a) The V1-V2 hypervariable regions (primers 402 and 403)

```
clone          1                                                  50
P1081827402    ..gttaaccc cactctgtgt tactttaaat tgcactgatt gggggaatgc
P1081828402    ......... .....TGTGT TACTTTAAAT TGCACTGATT GGGGGAATGC
P1081103402    .......ccc cactctgtgt tactttaaat tgcactgatt gggggaatgc
P1081107402    .......ccc ctctctgtgt tactttaaat tgcactgatt gggggaatgc
P1081006402    ......... ....ctgtgt tactttaaat tgcactgatt tggggaatgc
P1081007402    ...ttaaccc cactctgtgt tactttaaat tgcactgatt tggggaatgc
P1081501402    .......ccc cactctgtgt tactttaaat tgcactgatt tgaggaatgc
P1081504402    .......CCC CACTCTGTGT TACTTTAAAT TGCACTGATT TGAGGAATGC
P1081733402    AAATTAACCC CACTCTGTGT TACTTTAAAT TGCACTGATT TGGGGAATGC
P1081906402    ......... .........t tactttaaat tgcactgatt tgaggaatgc
P0120810402    .......ccc cactctgtgt tactttaaat tgcactgatt gggggaatgc
P0121604402    ......... ......... ......... ......... ...agaatgc
P0121614402    ......... ......... ......aaat tgcactgatt gggagaatgc
P0822407402    AAATTAACCC CACTCTGTGT TACTTTAAAT TGCACTGATG TGAGGAATGC
P1231410402    .AATTAACCC CACTCTGTGT TACTTTAAAT TGCACTGATG TGAGGAATGC
P1231411402    ......... ......... ......... .gcactgatg tgaggaatgc

               51                                                 100
P1081827402    tactaatacc aatag..... ......... .........t agtagtgggg
P1081828402    TACTAATACC AATAG..... ......... .........T AGTAGTGGGG
P1081103402    tactaatatc aataa..... ......... .........t agtagtgggg
P1081107402    tactaatatc aataa..... ......... .........t agtagtgggg
P1081006402    tactaatatc aataa..... ......... .........t agtagtgggg
P1081007402    tactaatatc aataa..... ......... .........t agtagtgggg
P1081501402    tactaatacc aataatgc.. .tactaat.c ca...atagt agtagttggg
P1081504402    TACTAATACC AATAATGC.. .TACTAATAC CA...ATAGT AGTAGTTGGG
P1081733402    TACTAATATC AATAA..... ......... .........T AGTAGTGGGG
P1081906402    tactaatacc aataatgc.. .tactaatac ca...atagt agtagttggg
P0120810402    tactaatatc aatag..... ......... .........t agtagtgggg
P0121604402    tactaatacc ...aatgc.. .tactaatac ca...atagt agtagtgggg
P0121614402    tactaatgct actaatacca atgctactaa taccaatagt agtagtgggg
P0822407402    TACTAATGCC AATAATGC.. .TACTAATAC CACTAGTAGT AGTATTGGGA
P1231410402    TACTAATGCC AATAATGC.. .TACTAATAC CAC...TAGT AGTATTGGGG
P1231411402    tactaatgcc aataatgc.. .tactaatac cac...tagt agtattgggg

               101                                                150
P1081827402    aaacgatgag aggagaaata aaaaactgct cTttcaatat caccacaagt
P1081828402    AAACGATGAG AGGAGAAATA AAAAACTGCT CTTTCAATAT CACCACAAGT
P1081103402    aaccgatgag aggagaaata aaaaactgct ctttcaatat caccacaagt
P1081107402    aaccgatgag aggagaaata aaaaactgct ctttcaatat caccacaagt
P1081006402    aaccgatgag aggagaaata aaaaactgct ctttcaatat caccacaagt
P1081007402    aaccgatgag aggagaaata aaaaactgct ctttcaatat caccacaagt
P1081501402    aaaagatgag aggagaaata aaaaactgct ctttcaatat caccacaagt
P1081504402    AAAAGATGAG AGGAGAAATA AAAAACTGCT CTTTCAATAT CACCACAAGT
P1081733402    AACCGATGAG AGGAGAAATA AAAAACTGCT CTTTCAATAT CACCACAAGT
P1081906402    aaCCGATGAG AGGAGAAATA AAAAACTGCT CTTTCAATAT CACCACAAGT
P0120810402    aaccgatgag aggagaaata aaaaactgct ctttcaatat caccacaagt
P0121604402    aagagatgag aggagaaata aaaaactgct ctttcaatat caccacaagt
P0121614402    aagagatgag aggagaaata aaaaacagct cttcaatat. ..........
P0822407402    AAGAGATGAG AGGAGAAATA AAAAACTGCT CTTTCAATAT CACCACAAGT
P1231410402    AAGAGATGAG AGGAGAAATA AAAAACTGCT CTTTCAATAT CACCACAAGT
P1231411402    aagagatgag aggagaaata aaaaactgct ctttcaatat caccacaagt

               151                                                200
P1081827402    ataagagata aggtgCagaa agaatatgca ctttttttata aacttgatgt
P1081828402    ATAAGAGATA AGGTGCAGAA AGAATATGCA CTTTTTTATA AACTTGATGT
P1081103402    ataagagata aggtgcaaaa agaatatgca ctttttttata aacttgatgt
P1081107402    ataagagata aggtgcaaaa agaatatgca ctttttttata aacttgatgt
```

```
P1081006402  ataagagata aggtgcagaa agaatatgca ctttttata aacttgatgt
P1081007402  ataagagata aggtgcagaa agaatatgca ctttttata aacttgatgt
P1081501402  ataagagata aggtgcagaa agaatatgca ctttttata a.........
P1081504402  ATAAGAGATA AGGTGCAGAA AGAATATGCA CTTTTTTATA ACTTGATGTA
P1081733402  ATAAGAGATA AGGTGCAGAA AGAATATGCA CTTTTTTATA AACTTGATGT
P1081906402  ATAAGAGATA AGGTgCAGAA AGAATATGCA CTTTTTTATA AACTTGATGT
P0120810402  ataag..... .......... .......... .......... ..........
P0121604402  ataagagata aggtgcagaa agaatatgca ctttttata aacttgatgt
P0121614402  .......... .......... .......... .......... ..........
P0822407402  ATAAGAGATA AGGTGCAGAA AGAATATGCA CTTTTTTata aacttgatgt
P1231410402  ATAAGAGATA AGGTTGCAGA AAGAATATCA CTTTTTTATA AACTTGATGT
P1231411402  ataagagata aggtgcagaa agaatatgca ctttttata aacttgatgt

             201                                                250
P1081827402  agtaccaa.. .......... .......... .......... ..........
P1081828402  AGTACCAATA GATGAGGATA ATacTAAT.. .......... ..........
P1081103402  agtaccaata gatgaggata atactaatac cagctatagg ttata.....
P1081107402  agtaccaata gatgaggata atactaatac cagctatagg ttgatagttg
P1081006402  agtaccaata gatgaggata atactaatac cagctatggt tgataag...
P1081007402  agtaccaata gatgaggata atactaatac cagctatagg ttg.......
P1081501402  .......... .......... .......... .......... ..........
P1081504402  .......... .......... .......... .......... ..........
P1081733402  AGTACCAATA GATGAGGATA ATACTAATAC CAGCCAGCTA TAGGTT....
P1081906402  AGTACCAATA GATGAGGATA ATACTAATAC CAGCTATAGG TTgATAAGTT
P0120810402  .......... .......... .......... .......... ..........
P0121604402  agtaccaata gatgaggata atactaatac cagctatagg ttgataagtt
P0121614402  .......... .......... .......... .......... ..........
P0822407402  agtaccaata gatgaggata atactaatac cagctatagg ttgataagtt
P1231410402  AGTATTAATA GATGAGGATA ATACTAATAC CACCGCTATA GGTTGATAA.
P1231411402  agtaccaata gatgaggata at........ .......... ..........

             251                        277
P1081827402  .......... .......... .......
P1081828402  .......... .......... .......
P1081103402  .......... .......... .......
P1081107402  tatcc..... .......... .......
P1081006402  .......... .......... .......
P1081007402  .......... .......... .......
P1081501402  .......... .......... .......
P1081504402  .......... .......... .......
P1081733402  .......... .......... .......
P1081906402  GTAATACCTC AGTCATTACA CAGG...
P0120810402  .......... .......... .......
P0121604402  gtaatacctc agtcattaca caggcct
P0121614402  .......... .......... .......
P0822407402  gtaatacctc agt....... .......
P1231410402  .......... .......... .......
P1231411402  .......... .......... .......

             1                                                  50
P1081828403  .......... .......... .......... .......... ..........
P1081501403  .......... .......... .......... .......... ..........
P1081504403  gtttaatagt acttggaatt attataatgg gtacttggaa ttcaacacaa
P0120810403  .......... .......... .......... .......... ..........
P0121604403  .......... .......... .......... .......... ..........
P1390610403  .......... .......... .......... .......... ..........
P1396b09403  .......... .......... .......... .......... ..........

             51                                                 100
P1081828403  .GGGATACCT TTGGACaggc ctgtgtaatg actgaggtat tacaatttat
P1081501403  .....tatct ttggccaggc ctgtgtaatg actgaggtat tacaacttat
P1081504403  cataatactg gagaaaatat cacactccca tgtagaataa aacaaattat
P0120810403  .......... .........c ctgtgtaatg actgaggtat tacaacttat
P0121604403  .......... .......... .......... .ctgaggtat tacaacttat
P1390610403  .......... .......... .......... .......... ..........
```

```
P1396b09403  ..........  ..........  ..........  .ctgaggtat tacaacttat

             101                                                    150
P1081828403  caacctatag ctggtattag tattatcctc atctattggt actacatcaa
P1081501403  caacctatag ctggtattag tattatcctc atctattggt actacatcaa
P1081504403  aaacatgtgg caggaagtag gaaaacgaat gtatgcccct cccatcagag
P0120810403  caacctatag ctggtattag tattatcctc atctattggt actacatcaa
P0121604403  caacctatag ctggtattag tattatcctc atctattggt actacatcaa
P1390610403  cctcatgagc taggtattag tattatcctc atctattggt actacatcaa
P1396b09403  caacctatag ctggtattag tattatcctc atctattggt actacatcaa

             151                                                    200
P1081828403  gtttataaaa aagtgcatat tctttctgca ccttatctct tatactt...
P1081501403  ctttataaaa aagtgcatat tctttctgca ccttatctct tatacttgtg
P1081504403  gacaaattag atgttcatca aatattacgg gctactatta acaagagatg
P0120810403  gtttataaaa aagtgcatat tctttctgca ccttatctct tatacttgtg
P0121604403  gtttataaaa aagtgcatat tctttctgca ccttatctct tatacttgtg
P1390610403  gtttataaaa aagtgcatat tctttctgca ccttatctct tatacttgtg
P1396b09403  gtttataaaa aagtgcatat tctttctgca ccttatctct tatacttgtg

             201                                                    248
P1081828403  ..........  ..........  ..........  ..........  ........
P1081501403  gtgatattga aagagcagtt ttttatttct cctctcatc.  ........
P1081504403  gtggtaacaa gaac......  ..........  ..........  ........
P0120810403  gtgatattga aagagcagtt ttttatttct cctctcatcg gtttcccc
P0121604403  gtgatattga aagagcagtt ttttatttct cctctcatct cttcccc.
P1390610403  gtgatattga aagagcagtt ttttatttct cctctcatct ctgcccc.
P1396b09403  gtgacattga aagagcagtt ttttatttct cctctcatct ctgcccc.
```

b) The V3 hypervariable region (primer 307 antisense)

```
clone        1                                                      50
P1081827307  .......... ......gttt tattcccaaa ttgttctctt aattttataa
P1081828307  .........g actattgttt tattcccaaa ttgttctctt aattttataa
P1081103307  .......... .......... .......... .......... ..........
P1081107307  .......... .......... .......... ..gttctcct aattttataa
P1081006307  .......... ......gttc tattcccaaa ttgttctcct aattttataa
P1081007307  .......... .......... .......... .......... ..........
P1081501307  .......... .......... .......... ..GTTCTCCT AATTTTATAA
P1081904307  .........g actattgttt tattctcaaa ttgttctctt aattttataa
P1081906307  .......... .......TTT TATTCTCAAA TTGTTCTCTT AATTTTATAA
P0120810307  .......... .......... ....cccaaa ttgttctcct aattttataa
P0121603307  .......... .......... .......... .......... ..........
P0121604307  .......... ......gctt tatgcccaaa ttgttctctt aattttataa
P0822407307  TTACTTAAAG ACTATTGTTT TATTCTCAAA TTGTTCTCTT AATTTTATAA
P1231409307  .......AAG ACTATTGTTT TATTCTCAAA TTGTTCTCTT AATTTTATAA
P1231410307  .......AAG ACTATTGTTT TATTCTCAAA TTGTTCTCTT AATTTTATAA
P1231411307  .......AAG ACTATTGTTT TATTCTCAAA TTGTTCTCTT AATTTTATAA
P1390516307  .......... ......gttt tatTctcaaa ttgttctctt aattttataa
P1390520307  .......... ......GTTT TATTCTCAAA TTGTTCTCTT AATTTTATAA
P1390501307  .......... .....tgttt tattctcaaa ttgttctctt aattttataa
P1390503307  .......... .....tgttt tattctcaaa ttgttctctt aattttataa
P1390608307  .......... ......gttt tattctcaaa ttgttctctt aattttataa
P1390610307  .......... .......... ....CTCAAA TTGTTCTCTT AATTTTATAA
P1390611307  .......... .......... .......... .......... ..........
P1396b09307  .......... .......... ....CCCAAA TTGTTCTCTT AATTTTGTAA
P1396b12307  ......gact attgttttat tattcccaaa ttgttctctt aattttgtaa

             51                                                     100
P1081827307  ctatctgtct taaagtttca ttccattttg ctctactaag gttacaatgt
P1081828307  ctatctgtct taaagtttca ttccattttg ctctactaag gttacaatgt
P1081103307  ctatctgtct taaagtttca ttccattttg ctctactaag gttacaatgt
P1081107307  ctatctgtct taaagtttca ttccattttg ctctactaag gttacaatgt
```

```
P1081006307  ctatctgttt taaagtttca ttccattttg ctctactaag gttacaatgt
P1081007307  ctatctgttt taaagtttca ttccattttg ctctactaag gttacaatgt
P1081501307  CTATCTGTCT TAAAGTTTCA TTCCATTTTG CTCTACTAAG GTTACAATGT
P1081904307  ctatctgttt taaagtttca ttccattttg ctctactaat gttacaatgc
P1081906307  CTATCTGTTT TAAAGTTTCA TTCCATTTTG CTCTACTAAT GTTACAATGC
P0120810307  ctatCTGTCT TAAAGTTTCA TTCCATtTTG CTCTACTAAG GTTACAATGT
P0121603307  ........TT TTAAGTTTCA TTCCATTTTG CTCTACTAAG GTTACAATGT
P0121604307  ctatctgttt taaagtttca ttccattttg ctctactaag gttacgatgt
P0822407307  CTATCTGTTT TAAAGTTTCA TTCCATTTTG CTCTACTAAT GTTACAATGT
P1231409307  CTATCTGTTT TAAAGTTTCA TTCCATTTTG CTCTACTAAT GTTACAATGT
P1231410307  CTATCTGTTT TAAAGTTTCA TTCCATTTTG CTCTACTAAT GTTACAATGT
P1231411307  CTATCTGTTT TAAAGTTTCA TTCCATTTTG CTCTACTAAT GTTACAATGT
P1390516307  ctatctgttt taaagtttca ttccattttg ctctactaat gttacAatgt
P1390520307  CTATCTGTTT TAAAGTTTCA TTCCATTTTG CTCTACTAAT GTTACAATGT
P1390501307  ctatctgttt taaagtttca ttccattttg ctctactaat gttacaatgt
P1390503307  ctatctgttt taaagtttca ttccattttg ctctactaat gttacaatgt
P1390608307  ctatctgttt taaagtttca ttccattttg ctctactaat gttacaatgt
P1390610307  CTATCTGTTT TAAAGTTTCA TTCCATTTTG CTCTACTAAT GTTACAATGT
P1390611307  CTATCTGTTT TAAAGTTTCA TTCCATTTTG CTCTACTAAT GTTACAATGT
P1396b09307  CTATCTGTTT TAAAGTTTCA TTCCATTTTG CTCTACTAAT GTTACAATGT
P1396b12307  ctatctgttt taaagtttca ttccattttg ctctactaat gttacaatgt

             101                                                150
P1081827307  gcttgtctta tatctcctat tatttctcct gttgtataaa atgctctccc
P1081828307  gcttgtctta tatctcctat tatttctcct gttgtataaa atgctctccc
P1081103307  gcttgtctta tatctcctat tatatctcct gttgtataaa atgctctccc
P1081107307  gcttgtctta tatctcctat tatatctcct gttgtataaa atgctctccc
P1081006307  gcttgtctta tatctcctat tatttgtcct gttgtataaa atgctccccc
P1081007307  gcttgtctta tatctcctat tatttgtcct gttgtataaa atgctccccc
P1081501307  GCTTGTCTTA TATCTCCTAT TATTTGTCCT GTTGTATAAA ATGCTCTCCC
P1081904307  gcttgtctta tatctcctat tattcctcct gttgtataaa atgctctccc
P1081906307  GCTTGTCTTA TATCTCCTAT TATTCCTCCT GTTGTATAAA ATGCTCTCCC
P0120810307  GCTTGTCTTA TATCTCCTAT TATGTCTCCT GTTGTATAAA ATGCTCTCCC
P0121603307  GCTTGTCTTA TATCTCCTAT TATGTCTCCT GTTGCATAAA ATGCACTCCC
P0121604307  gcttgtctta tatctcctat tatgtctcct gttgcataaa atgcactccc
P0822407307  GCTTGTCTTA TATTTCCTAT TATTTGTTCT GTTGTATAAA CTGCTCTCCC
P1231409307  GCTTGTCTTA TATTTCCTAT TATTTGTTCT GTTGTATAAA CTGCTCTCCC
P1231410307  GCTTGTCTTA TATTTCCTAT TATTTGTTCT GTTGTATAAA CTGCTCTCCC
P1231411307  GCTTGTCTTA TATTTCCTAT TATTTGTTCT GTTGTATAAA CTGCTCTCCC
P1390516307  gcttgtctta tatctcctat tatTcctcct gttgcataaa atgcactccc
P1390520307  GCTTGTCTTA TATCTCCTAT TATTCCTCCT GTTGCATAAA ATGCACTCCC
P1390501307  gcttgtctta tatctcctat tattcctcct gttgcataaa atgcactccc
P1390503307  gcttgtctta tatctcctat tattcctcct gttgcataaa atgcactccc
P1390608307  gcttgtctta tatctcctat tattcctcct gttgcataaa atgcactccc
P1390610307  GCTTGTCTTA TATCTCCTAT TATTCCTCCT GTTGCATAAA ATGCACTCCC
P1390611307  GCTTGTCTTA TATCTCCTAT TATTCCTCCT GTTGCATAAA ATGCACTCCC
P1396b09307  GCTTGCCTTA TATCTCCTAT TATGTCTCCT GTTGCATAAA ATGCACTCCC
P1396b12307  gcttgcctta tatctcctat tatgtctcct gttgcataaa atgcactccc

             151                                                200
P1081827307  tggtcctata tgtatacttt ttcttgtatt gttgttggg. ..........
P1081828307  tggtcctata tgtatacttt ttcttgtatt gttgttgggt cttgtacaat
P1081103307  tggtcctata tgtatacttt ttcttgtatt gtt.ttggtc tgtacatatt
P1081107307  tggtcctata tgtatacttt ttcttgtatt gttgttgggc tgtacattat
P1081006307  tggtcctata tgtatacttt ttcttgtatt gttgttgggt cttgtacaat
P1081007307  tggtcctata tgtatacttt ttcttgtatt gttgttgggt cttgtacaat
P1081501307  TGGTCCTATA GGTATACTTT TTCTTGTATT GTTGTTGGGT CTTGTACAAT
P1081904307  tggtcctata tgtatacttt ttcttgtatt gttgttgggt cttgtacaat
P1081906307  TGGTCCTATA TGTATACTTT TTCTTGTATT GTTGTTGGGT CTTGTACAAT
P0120810307  TGGTCCTATA GGTATACTTT TTCTTGTGTT GTTGTTGGGT CTTgTACAAT
P0121603307  TGGTCCTATA TGTATACCTT TTCTTGTATT GTTGTTGGGT CTTGTACAAT
P0121604307  tggtcctata tgtatacctt ttcttgtatt gttgttgggt cttgtacaat
P0822407307  TGGTCCTATA TATATCCTTT TTTT...... .......... ..........
P1231409307  TGGTCCTATA TGTATCCTTT TTCTTGTATT GTTGTTGGGT CTTGTACAAT
P1231410307  TGGTCCTATA TGTATCCTTT TTCTTGTATT GTTGTTGGGT CTTGTACAAT
P1231411307  TGGTCCTATA TGTATCCTTT TTCTTGTATT GTTGTTGGGT CTTGTACAAT
P1390516307  tggtcctata tgtatacctt ttcttgtatt gttgttgggt cttgtacaat
P1390520307  TGGTCCTATA TGTATACCTT TTCTTGTATT GTTGTTGGGT CTTGTACAAT
```

```
P1390501307  tggtcctata tgtatacctt ttcttgtatt gttgttgggt cttgtacaat
P1390503307  tggtcctata tgtatacctt ttcttgtatt gttgttgggt cttgtacaat
P1390608307  tggtcctata tgtatacctt ttcttgtatt gttgttgggt cttgtacaat
P1390610307  TGGTCCTATA TGTATACCTT TTCTTGTATT GTTGTTGGGT CTTGTACAAT
P1390611307  TGGTCCTATA TGTATACCTT TTCTTGTATT GTTGTTGGGT CTTGTACAAT
P1396b09307  TGGTCCTATA TGTATACCTT TTCTTGTATT GTTGTTGGGT CTTGTACAAT
P1396b12307  tggtcctata tgtatacctt ttcttgtatt gttgttgggt cttgtacaat

             201                                                 250
P1081827307  .......... .......... .......... .......... ..........
P1081828307  taatttctac agattccttc agctgtacta ttatt..... ..........
P1081103307  ctc....... .......... .......... .......... ..........
P1081107307  tctc...... .......... .......... .......... ..........
P1081006307  taattct... .......... .......... .......... ..........
P1081007307  taattct... .......... .......... .......... ..........
P1081501307  TAATTACTAC AGATTCCTTC AGCTGTACTA TTATGGTTTT AGCATT....
P1081904307  taattactac agattcattc agctgtac.. .......... ..........
P1081906307  TAATTACTAC AGATTCATTC AGCTGTACTA TTATgGTTTT ..........
P0120810307  TAATTTCTAC AGATTCCTTC AGCTGtacta ttatggtttt ..........
P0121603307  TAATTACTAC AGATTCATTC AGCTGTACTA TTATGGTTT. ..........
P0121604307  taattactac agattcatt. .......... .......... ..........
P0822407307  .......... .......... .......... .......... ..........
P1231409307  TAATTACTAC AGATTCATTC AGCTGTACTA TTATG..... ..........
P1231410307  TAA....... .......... .......... .......... ..........
P1231411307  TAATTACTAC AGA....... .......... .......... ..........
P1390516307  taattactac agattcattc agctgtacta tatgtttt.. ..........
P1390520307  TAATTACTAC AGATTCATTC AGCTGTACTA TATGTTTTAG CAT.......
P1390501307  taattactac agattcattc agctgtacta tatgttttag ..........
P1390503307  taattactac agattcattc agctgtacta tatgttt... ..........
P1390608307  taattactac agattcattc agtgtactat tatggtttta gcattgtctg
P1390610307  TAATTACTAC AGATTCATTC AGCTGT.... .......... ..........
P1390611307  TAATTACTAC AGATTCATTC AGCG...... .......... ..........
P1396b09307  TAATTACTAC AGATTCATTC AGCTGTACTA TTATGGTTTT AGCATTGTCT
P1396b12307  taattacta. .......... .......... .......... ..........

             251                         279
P1081827307  .......... .......... .........
P1081828307  .......... .......... .........
P1081103307  .......... .......... .........
P1081107307  .......... .......... .........
P1081006307  .......... .......... .........
P1081007307  .......... .......... .........
P1081501307  .......... .......... .........
P1081904307  .......... .......... .........
P1081906307  .......... .......... .........
P0120810307  .......... .......... .........
P0121603307  .......... .......... .........
P0121604307  .......... .......... .........
P0822407307  .......... .......... .........
P1231409307  .......... .......... .........
P1231410307  .......... .......... .........
P1231411307  .......... .......... .........
P1390516307  .......... .......... .........
P1390520307  .......... .......... .........
P1390501307  .......... .......... .........
P1390503307  .......... .......... .........
P1390608307  tgaa...... .......... .........
P1390610307  .......... .......... .........
P1390611307  .......... .......... .........
P1396b09307  GTGAAATTGC TAGATCTAAT TACTACCCC
P1396b12307  .......... .......... .........
```

c) V4-C4-V5 regions (primers 407 sense)

```
clone          1                                                  50
P1081827407    .......... .......... .....gtact tggaattcac cacaacttaa
P1081828407    .......... .......GTT TAATAGTACT TGGAATTCAC CACAACTTAA
P1081103407    .......... .......... ..CTTGGAAT TATTCTAATG ATACTTGGAG
P1081107407    .......... .......... .......... .......... ...cttggag
P1081004407    .......... .......... .....ggaat tattctaatg gtacttggat
P1081006407    .......... .........g tacttggaat tattctaatg gtaattggag
P1081007407    .......... .........g tacttggaat tattctaatg gtaattggag
P1081504407    .......... .GTTTAATAG TACTTGGAAT TATTATAATG GTACTTGGAA
P1081904407    .......... .......... .....ggaat tattctaatg gtacttggat
P1081906407    .........C TGTTTAATAG TACTTGGAAT TATTCTAATG GTACTTGGAT
P0120810407    .......... .........G TACTTGGATC TATTCTAATG GTACTTGGAA
P0121603407    .......... .......... .......... .......... ..........
P0121604407    ...CTGTTTA ATAGTACTTG GAATTCAACA TGGGATTTAA CACAACTTAA
P0121614407    CAACTGTTTA ATAGTACTTG GAATTCAACA TGGGATTTAA CACAACTTAA
P0822407407    CAACTGTAAA ATAGTACTTG GAATTCAACA TGGGATTTAA CACAACTTAA
P1231409407    .......... ......CTTG GAATTCAACA TGGGATTTAA CACAACTTAA
P1390501407    .......... .......... .......... ........aa cacaacttaa
P1390516407    ........TA ATAGtACTTG GAATTCAACA TGGGATTTAA CACAACTTAA
P1390520407    .......... ......CTTG GAATTCAACA TGGGATTTAA CACAACTTAA
P1390503407    ........TA ATAGtACTTG GAATTCAACA TGGGATTTAA CACAACTTAA
P1390603407    .......... .......... .......... .......TAA CACAACTTAA
P1390604407    .......... .......... .......... ........AA CACAACTTAA
P1390606407    .......... .......... .......... ........AA CACAACTTAA
P1390608407    .......... ......CTTG GAATTCAACA TGGGATTTAA CACAACTTAA
P1390610407    .....GTTTA ATAGTACTTG GAATTCAACA TGGGATTTAA CACAACTTAA
P1390611407    .......... ......CTTG GAATTCAACA TGGGATTTAA CACAACTTAA
P1396b07407    .......... .......... .......... ........aa cacaattt..
P1396b09407    .......... ...GAACTTG GAATTCAACA TGGGATTTAA CACAATTT..
P1396b12407    .......... .......... .......... .gggatttaa cacaattt..
P1396b12407    .......... ...GTACTTG GAATTCAACA TGGGATTTAA CACAATTT..

               51                                                 100
P1081827407    taatacttgg aataatactg aagaaaatat cacactccca tgtagaataa
P1081828407    TAATACTTGG AATAATACTG AAGAAAATAT CACACTCCCA TGTAGAATAA
P1081103407    TTCAACACAA CATAATACTG AAGAAAATAT CACACTCCCA TGTAGAATAA
P1081107407    ttcaacacaa cataatactg aagaaaatat cacactccca tgtagaataa
P1081004407    ttcaacacaa cataatactg aataaaatat cacactccca tgtagaataa
P1081006407    TTCAACACAA CATAATACTG AAGAAAATAT CACACTCCCA TGTAGAATAA
P1081007407    ttcaacacaa cataatactg aagaaaatat cacactccca tgtagaataa
P1081504407    TTCAACACAA CATAATACTG GAGAAAATAT CACACTCCCA TGTAGAATAA
P1081904407    ttcaacacaa cataatactg aagaaaatat cacactccca tgtagaataa
P1081906407    TTCAACACAA CATAATACTG AAGAAAATAT CACACTCCCA TGTAGAATAA
P0120810407    TCCAACACAA CATAATACTG AAGAAAATAT CACACTCCCA TGTAGAATAA
P0121603407    .......... ..TAA...AG AAGAAAATAT CACACTCCCA TGTAGAATAA
P0121604407    TAGTACTCGA AATAA...AG AAGAAAATAT CACACTCCCA TGTAGAATAA
P0121614407    TAGTACTCGG AATAA...AG AAGAAAATAT CACACTCCCA TGTAGAATAA
P0822407407    TAGTACTCAG AATAA...AG AAGAAAATAT CACACTCCCA TGTAGAATAA
P1231409407    TAGTACTCAG GATAA...AG AAGAAAATAT CACACTCCCA TGTAGAATAA
P1390501407    tagtactcag aataa...ag aagaaaatat cacactccca tgtagaataa
P1390516407    TAGTACTCAG AATAA...AG AAGAAAATAT CACACTCCCA TGTAGAATAA
P1390520407    TAGTACTCAG AATAA...AG AAGAAAATAT CACACTCCCA TGTAGAATAA
P1390503407    TAGTACTCAG AATAA...AG AAGAAAATAT CACACTCCCA TGTAGAATAA
P1390603407    TAGTACCCAG AATAA...AG AAGAAAATAT CACACTCCCA TGTAGAATAA
P1390604407    TAGTACCCAG AATAA...AG AAGAAAATAT CACACTCCCA TGTAGAATAA
P1390606407    TAGTACCCAG AATAA...AG AAGAAAATAT CACACTCCCA TGTAGAATAA
P1390608407    TAGTACTCAG AATAA...AG AAGAAAATAT CACACTCCCA TGTAGAATAA
P1390610407    TAGTACTCAG AATAA...AG AAGAAAATAT CACACTCCCA TGTAGAATAA
P1390611407    TAGTACTCAG AATAA...AG AAGAAAATAT CACACTCCCA TGTAGAATAA
P1396b07407    .......cag aatga...ag aagaaaatat cacactCCCA TGTAGAATAA
P1396b09407    .......... CAGAATGAAG AAGAAAATAT CACACTCCCA TGTAGAATAA
P1396b12407    .......cag aatga...ag aagaaaatat cacactccca tgtagaataa
P1396b12407    .......... CAGAATGAAG AAGAAAATAT CACACTCCCA TGTAGAATAA
```

```
                101                                                 150
P1081827407  aacaaattat aaacatgtgg caggaagtag gaaaagcaat gtatgcccct
P1081828407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P1081103407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCC
P1081107407  aacaaattat aaacatgtgg caggaagtag gaaaagcaat gtatgccccc
P1081004407  aacaaattat aaacatgtgg caggaagtag gaaaagcaat gtatgcccct
P1081006407  AACAAATTAT AAACATGTGG CAGGGAGTAG GAAAAGCAAT GTATGCCCCT
P1081007407  aacaaattat aaacatgtgg cagggagtag gaaaagcaat gtatgcccct
P1081504407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P1081904407  aacaaattat aaacatgtgg caggaagtag gaaaagcaat gtatgcccct
P1081906407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P0120810407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P0121603407  AACAAATTAT AAATATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P0121604407  AACAAATTAT AAATATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P0121614407  AACAAATTAT AAATATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P0822407407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGcAAt gtatgcccct
P1231409407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P1390501407  aacaaattat aaacatgtgg caggaagtag gaaaagcaat gtatgcccct
P1390516407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P1390520407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P1390503407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P1390603407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P1390604407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P1390606407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P1390608407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P1390610407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P1390611407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P1396b07407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P1396b09407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAAGCAAT GTATGCCCCT
P1396b12407  aacaaattat aaacatgtgg caggaagtag gaaaagcaat gtatgcccct
P1396b12407  AACAAATTAT AAACATGTGG CAGGAAGTAG GAAAA..... ..........

                151                                                 200
P1081827407  cccatcagag gacaaattag atgttcatca aatattacag ggctactatt
P1081828407  CCCATCAGAG GACAAATTAG ATGTTCATCA AATATTACAG GGCTACTATT
P1081103407  CCCATCAGAG GACAAATTAG ATGTTCATCa AATATTACAG GGCTACTATT
P1081107407  cccatcagag gacaaattag atgttcatca aatattacag ggctactatt
P1081004407  cccatcagag gacaaattag atgttcatca aatattacag ggctactatt
P1081006407  CCCATCAGAG GACAAATTAG ATGTTCATCA AATATTACAG GGCTACTATT
P1081007407  cccatcagag gacaaattag atgttcatca aatattacag ggctactatt
P1081504407  CCCATCAGAG GACAAATTAG ATGTTCATCA AATATTACAG GGCTACTATT
P1081904407  cccatcagag gacaaattag atgttcatca aatattacag ggctactatt
P1081906407  CCCATCAGAG GACAAATTAG ATGTTCATCA AATATTACAG GGCTACTATT
P0120810407  CCCATCAGAG GACAAATTAG ATGTTCATCA AATATTACAG GGCTACTATT
P0121603407  CCCATCAGAG GACAAATTAG ATGTTCATCA AATATTACag GGCTATTGTT
P0121604407  CCCATCAGAG GACAAATTAG ATGTTCATCA AATATTACAG GGCTATTATT
P0121614407  CCCATCAGAG GACAAATTAG ATGTTCATCA AATATTACAG GGCTATTATT
P0822407407  cccatcagcg gacaaattag atgttcatca aatattcaag ggctactatt
P1231409407  CCCATCAAAG GACAAATTAG ATGTTCATCA AATATTACAG GGCTACTATT
P1390501407  cccatcagag gacaaattag atgttcatca aatattacag ggctattatt
P1390516407  CCCATCAGAG GACAAATTAG ATGTTCATCA AATATTACAG GGCTATTATT
P1390520407  CCCATCAGAG GACAAATTAG ATGTtCATCA AATATTACAG GGCTATTATT
P1390503407  CCCATCAGAG GACAAATTAG ATGTTCATCA AATATTACAG GGCTATTATT
P1390603407  CCCATCAGAG GACAAATTAG ATGTTCATCA AATATTACAG GGCTATTATT
P1390604407  CCCATCAAAG GACAAATTAG ATGTTCATCA ACTATTACAG GGCTACTATT
P1390606407  CCCATCAGAG GACAAATTAG ATGTTCATCA AATATTACAG GGCTATTATT
P1390608407  CCCATCAAAG GACAAATTAG ATGTTCATCA ACTATTACAG GGCTACTATT
P1390610407  CCCATCAAAG GACAAATTAG ATGTTCATCA ACTATTACAG GGCTACTATT
P1390611407  CCCATCAAAG GACAAATTAG ATGTTCATCA ACTATTACAG GGCTACTATT
P1396b07407  CCCATCAGAG GACAAATTAG aTGTTCATCA AATATTACAG GGCTACTATT
P1396b09407  CCCATCAGAG GACAAATTAG ATGTTCATCA AATATTACAG GGCTACTATT
P1396b12407  cccatcagag gacaaattag atgttcatca aatattacag ggctactatt
P1396b12407  .......... .......... .......... .......... ..........

                201                                                 250
P1081827407  aacaagagat ggtggtaaca cg........ .......... ..........
P1081828407  AACAAGAGAT GGTGGT.... .......... .......... ..........
P1081103407  .......... .......... .......... .......... ..........
```

```
P1081107407  aac..... ..... ..... ..... .....
P1081004407  aacaagagat ggtggtaatg gtggtaacaa aagcaacg.. .....
P1081006407  AACAAGAGAT GGTGGTAACA ACACGAACGG GACCGAGATC TTCAGACCTG
P1081007407  aacaagagat ggtggt.... ..... ..... ..... .....
P1081504407  AACAAGAGAT GGTGGT.... ..... ..... ..... .....
P1081904407  aacaagagat ggtggtaatg gtggtaacaa aagcaacg.. .....
P1081906407  AACAAGAGAT GGTGGTAATG GTGGTAACAA AAGCAACGAC ACCACCGAGA
P0120810407  AACAAGAGAT GGTGGTAACA AGAGCGAGAC CGACCCC... .....
P0121603407  AACAAGAGAT GGTGGTAATG GTGGTA.... ..... .....
P0121604407  AACAAGAGAT GGTGGTAATG GTGGTAACAA A..... .....
P0121614407  AACAAGAGAT GGTGGTAATG GTGGTAACAA A..... .....
P0822407407  aacaagagat ggtggtaata gtggtaacaa aagcaacgac accaccgaga
P1231409407  AACAAGAGAT GGTGGTAATA GTGGTAACAA AAGCAACG.. .....
P1390501407  aacaagagat ggtggtaatg gtggtaacaa aa..... .....
P1390516407  AACAAGAGAT GGTGGTAATG GTGGTAACAA AAGCAACGAC ACCACCGAGA
P1390520407  AACAAGAGAT ..... ..... ..... .....
P1390503407  AACAAGAGAT GGTGGTAATG GTGGTAACAA AAGCAACGAC ACCACCGAGA
P1390603407  AACAAGAGAT GGTGGTAATG GTGGTAACAA AA..... .....
P1390604407  AACAAGAGAT GGTGGTAATA GTGGTAACAA AA..... .....
P1390606407  AACAAGAGAT GGTGGTAATG GTGGAACAAA AGCAAC.... .....
P1390608407  AACAAGAGAT GGTGGTAATA GTGGTAACAA AAGCAACGAC ACCACC....
P1390610407  AACAAGAGAT GGTGGTAATA GTGGTAACAA AAGCAACGAC ACCACCGAGA
P1390611407  AACAAGAGAT GGTGGTAATA GTGGTAACAA AAGCAACGAC ACCACCGAGA
P1396b07407  AACAAGAGAt GGTGGTGACA CGAGC..... ..... .....
P1396b09407  AACAAGAGAT GGTGGTGACA C..... ..... ..... .....
P1396b12407  aacaagagat ggtggtgaca cgag..... ..... .....
P1396b12407  ..... ..... ..... ..... .....

             251                                          300
P1081827407  ..... ..... ..... ..... .....
P1081828407  ..... ..... ..... ..... .....
P1081103407  ..... ..... ..... ..... .....
P1081107407  ..... ..... ..... ..... .....
P1081004407  ..... ..... ..... ..... .....
P1081006407  GGGGAGGAGA TATGAGGGAC AATTGGAGAA GTGAATTATA TAAATATAaA
P1081007407  ..... ..... ..... ..... .....
P1081504407  ..... ..... ..... ..... .....
P1081904407  ..... ..... ..... ..... .....
P1081906407  CCTTCAGACC TGGGGG.... ..... ..... .....
P0120810407  ..... ..... ..... ..... .....
P0121603407  ..... ..... ..... ..... .....
P0121604407  ..... ..... ..... ..... .....
P0121614407  ..... ..... ..... ..... .....
P0822407407  ccttcagacc tgggg..... ..... ..... .....
P1231409407  ..... ..... ..... ..... .....
P1390501407  ..... ..... ..... ..... .....
P1390516407  CCTTCAGACC TGGGGGAGGA GATATGA... ..... .....
P1390520407  ..... ..... ..... ..... .....
P1390503407  CCTTCAGACC TGGGGGAGGA GATATGA... ..... .....
P1390603407  ..... ..... ..... ..... .....
P1390604407  ..... ..... ..... ..... .....
P1390606407  ..... ..... ..... ..... .....
P1390608407  ..... ..... ..... ..... .....
P1390610407  TCTT..... ..... ..... ..... .....
P1390611407  TCTTCAGACC TGGGGGGAGG AGATATGAGG GACAATTGGA GAAGTGAATT
P1396b07407  ..... ..... ..... ..... .....
P1396b09407  ..... ..... ..... ..... .....
P1396b12407  ..... ..... ..... ..... .....
P1396b12407  ..... ..... ..... ..... .....

             301                              334
P1081827407  ..... ..... ..... ....
P1081828407  ..... ..... ..... ....
P1081103407  ..... ..... ..... ....
P1081107407  ..... ..... ..... ....
P1081004407  ..... ..... ..... ....
P1081006407  GTAGTAAAAa TTGAACCATT AGGAGTAGCA CCAC
P1081007407  ..... ..... ..... ....
```

```
P1081504407  .......... .......... .......... ....
P1081904407  .......... .......... .......... ....
P1081906407  .......... .......... .......... ....
P0120810407  .......... .......... .......... ....
P0121603407  .......... .......... .......... ....
P0121604407  .......... .......... .......... ....
P0121614407  .......... .......... .......... ....
P0822407407  .......... .......... .......... ....
P1231409407  .......... .......... .......... ....
P1390501407  .......... .......... .......... ....
P1390516407  .......... .......... .......... ....
P1390520407  .......... .......... .......... ....
P1390503407  .......... .......... .......... ....
P1390603407  .......... .......... .......... ....
P1390604407  .......... .......... .......... ....
P1390606407  .......... .......... .......... ....
P1390608407  .......... .......... .......... ....
P1390611407  .......... .......... .......... ....
P1396b07407  .......... .......... .......... ....
P1396b09407  .......... .......... .......... ....
P1396b12407  .......... .......... .......... ....
P1396b12407  .......... .......... .......... ....
```

d) gp41 (primer 367 antisense)

```
clone        1                                                  50
P1081827367  .......... .CCAAGGCAC AGTAGTGGTG CAAATGAGTT TTCCAGAGCA
P1081828367  .......... .ccaaggcac agtagtggtg caaatgagtt ttccagagca
P1081107367  ...CTAGCAT TCCAAGGCAC AGCAGTGGTG CAAATGAGTT TTCCAGAGCA
P1081501367  .......... .....ggcac agcagtggtg caaatgagtt ttccagagca
P1081504367  ......GCAT TCCAAGGCAC AGCAGTGGTG CAAATGAGTT TTCCAGAGCA
P1081906367  .......... .....ggcac agcagtggtg caaatgagtt ttccagagca
P0120810367  .......... .ccaaggcac agcagtggtg caaatgagtt ttccagagca
P0121603367  ......GCAT TCCAAGGCAC AGCAGTGGTG CAAATGAGTT TTCCAGAGCA
P0121604367  ggcttagcat tccaaggcac agcagtggtg caaatgagtt ttccagagca
P0121614367  .......... .......... .......... .......... TTCCAGAGCA
P1231409367  .......GTA TCCAAGGCAC AGCAGTGGTG CAAATGAGTT TTCCAGAGCA
P1390516367  .......GTA TCCAAGGCAC AGCAGTGGTG CAAATGAGTT TTCCAGAGCA
P1390520367  .......GTA TCCAAGGCAC AGCAGTGGTG CAAATGAGTT TTCCAGAGCA
P1390501367  .......GTA TCCAAGGCAC AGCAGTGGTG CAAATGAGTT TTCCAGAGCA
P1390503367  .......GTA TCCAAGGCAC AGCAGTGGTG CAAATGAGTT TTCCAGAGCA
P1390608367  ......GCAT TCCAAGGCAC AGCAGTGGTG CAAATGAGTT TTCCAGAGCA
P1390610367  ......GCAT TCCAAGGCAC AGCAGTGGTG CAAATGAGTT TTCCAGAGCA
P1390611367  ......GCAT TCCAAGGCAC AGCAGTGGTG CAAATGAGTT TTCCAGAGCA
P1396b09367  .......... .....GGCAC AGCAGTGGTG CAAATGAGTT TTCCAGAGCA
P1396b12367  ......gtat tccaaggcac agcagtggtg caaatgagtt ttccagagca

             51                                                 100
P1081827367  ACCCCAAATC CCTAGGAGCT GTTGATCCTT TAGGTATCTT TCCACAGCCA
P1081828367  accccaaatc cctaggagct gttgatcctt taggtatctt tccacagcca
P1081107367  ACCCCAAATC CCTAGGAGCT GTTGATCCTT TAGGTATCTT TCCACAGCCA
P1081501367  accccaaatc cctaggagct gttgatcctt taggtatctt tccacagcca
P1081504367  ACCCCAAATC CCTAGGAGCT GTTGATCCTT TAGGTATCTT TCCACAGCCA
P1081906367  accccaaatc cctaggagct gttgatcctt taggtatctt tccacagcca
P0120810367  accccaaatc cctaggagct gttgatcctt taggtatctt tccacagcca
P0121603367  ACCCCAAATC CCTAGGAGCT GTTGATCCTT TAGGTATCTT TCCACAGCCA
P0121604367  accccaaatc cctaggagct gttgatcctt taggtatctt tccacagcca
P0121614367  ACCCCAAATC CCTAGGAGCT GTTGATCCTT TAGGTATCTT TCCACAGCCA
P1231409367  ACCCCAAATC CCTAGGAGCT GTTGATCCTT TAGGTATCTT TCCACAGCCA
P1390516367  ACCCCAAATC CCTAGGAGCT GTTGATCCTT TAGGTATCTT TCCACaGCCA
P1390520367  ACCCCAAATC CCTAGGAGCT GTTGATCCTT TAGGTATCTT TCCACAGCCA
P1390501367  ACCCCAAATC CCTAGGAGCT GTTGATCCTT TAGGTATCTT TCCACaGCCA
P1390503367  ACCCCAAATC CCTAGGAGCT GTTGATCCTT TAGGTATCTT TCCACAGCCA
P1390608367  ACCCCAAATC CCTAGGAGCT GTTGATCCTT TAGGTATCTT TCCACAGCCA
```

```
P1390610367  ACCCCAAATC CCTAGGAGCT GTTGATCCTT TAGGTATCTT TCCACAGCCA
P1390611367  ACCCCAAATC CCTAGGAGCT GTTGATCCTT TAGGTATCTT TCCACAGCCA
P1396b09367  ACCCCAAATC CCTAGGAGCT GTTGATCCTT TAGGTATCTT TCCACAGCCA
P1396b12367  accccaaatc cctaggagct gttgatcctt taggtatctt tccacagcca

             101                                                150
P1081827367  GGACTCTTGC CTGGAGCTGC TTGATGCCCC AGACTGTGAG TTGCAACAGA
P1081828367  ggactcttgc ctggagctgc ttgatgcccc agactgtgag ttgcaacaga
P1081107367  GGACTCTTGC CTGGAGCTGC TTGATGCCCC AGACTGTGAG TTGCAACAGA
P1081501367  ggactcttgc ctggagctgc ttgatgcccc agactgtgag ttgcaacaga
P1081504367  GGACTCttGC CTGGAGCTGC TTGATGCCCC AGACTGTGAG TTGcAACAGA
P1081906367  ggactcttgc ctggagctgc ttgatgcccc agactgtgag ttgcaacaga
P0120810367  ggactcttgc ctggagctgc ttaatgcccc agactgtgag ttgcaacaga
P0121603367  GGACTCTTGC CTGGAGCTGC TTGATGCCCC AGACTGTGAG TTGCAACAGA
P0121604367  ggactcttgc ctggagctgc ttgatgcccc agactgtgag ttgcaacaga
P0121614367  GGACTCTTGC CTGGAGCTGC TTGATGCCCC AGACTGTGAG TTGCAACAGA
P1231409367  GGACTCTTGC CTGGAGCTGC TTGATGCCCC AGACTGTGAG TTGCAACAGA
P1390516367  GGACTCTTGC CTGGAGCTGC TTAATGCCCC AGACTGTGAG TTGCAACAGA
P1390520367  GGACTCTTGC CTGGAGCTGC TTAATGCCCC AGACTGTGAG TTGCAACAGA
P1390501367  GGACTCTTGC CTGGAGCTGC TTAATGCCCC AGACTGTGAG TTGCAACAGA
P1390503367  GGACTCTTGC CTGGAGCTGC TTAATGCCCC AGACTGTGAG TTGCAACAGA
P1390608367  GGACTCTTGC CTGGAGCTGC TTGATGCCCC AGACTGTGAG TTGCAACAGA
P1390610367  GGACTCTTGC CTGGAGCTGC TTGATgCCCC AGACTGTGAG TTGCAACAGA
P1390611367  GGACTCTTGC CTGGAGCTGC TTGATGCCCC AGACTGTGAG TTGCAACAGA
P1396b09367  GGACTCTTGC CTGGAGCTGT TTGATGCCCC AGACTGTGAG TTGCAACAAA
P1396b12367  ggactcttgc ctggagctgt ttgatgcccc agactgtgag ttgcaacaaa

             151                                                200
P1081827367  TGCTGTTGCG CCTCAATAGC CCTCAGCAGA TTGTTCTGCT GTTGCACTAT
P1081828367  tgctgttgcg cctcaatagc cctcagcaga ttgttctgct gttgcactat
P1081107367  TGCTGTTGCG CCTCAATAGC CCTCAGCAAA TTGTTCT... ..........
P1081501367  t......... .......... .......... .......... ..........
P1081504367  TGCTGTTgcG CCTCAATAGC CCTCAGCAAA TTGTTCTGCT GTTGCACTAT
P1081906367  tgctgttgcg cctcaatagc cctcagcaaa ttgttctgct gctgcactat
P0120810367  tgctgttgcg cctcaatagc cctcagcaaa ttgttctgct gttgcactat
P0121603367  TGCTGTTGCG CCT....... .......... .......... ..........
P0121604367  tgctgttgcg cctcaatagc cctcagcaaa ttgttctgct gttgc.....
P0121614367  TGCTGTTGCG CCTCAATAGC CCTCAGCAAA TTGTTCTGCT GTTGCACTAT
P1231409367  TGCTGTTGCG CCTCAATAGC CCTCAGCAAA TTGTTCTGCT GTTGCACTAT
P1390516367  TGCTGTTGCG CCTCAATAGC CCTCAGCAAA TTGTTCTGCT GTTGCACTAT
P1390520367  TGCTGTTGCG CCTCAATAGC CCTCAGCAAA TTGTTCTGCT GTTGCACTAT
P1390501367  TGCTGTTGCG CCTCAATAGC CCTCAGCAAA TTGTTCTGCT GTTGCACTAT
P1390503367  TGCTGTTGCG CCTCAATAGC CCTCAGCAAA TTGTTCTGCT GTTGCACTAT
P1390608367  TGCTGTTGCG CCTCAATAGC CCTCAGCAAA TTGTTCTGCT GTTGCACTAT
P1390610367  TGCTGTTgcG CCTCAATAGC CCTCAGCAAA TTGTTCTGCT GTTGCACTAT
P1390611367  TGCTGTTGCG CCTCAATAGC CCTCAGCAAA TTGTTCTGCT GTTGCACTAT
P1396b09367  TGCTGTTGCG CCTCAATAGC CCTCAGCAAA TTGTTCTGCT GTTGCACTAT
P1396b12367  tgctgttgcg cctcaatagc cctcagcaaa ttgttctgct gttgcactat

             201                                                250
P1081827367  ACCAGACAAT AATAGTCTGG CCTGTACCGT CAGCGCCATT GACGCTGCGC
P1081828367  accagacaat aatagtctgg cc........ .......... ..........
P1081107367  .......... .......... .......... .......... ..........
P1081501367  .......... .......... .......... .......... ..........
P1081504367  ACCAGACAAT AATAG..... .......... .......... ..........
P1081906367  accagacaat aatagtctgg cctgt..... .......... ..........
P0120810367  accagacaat aatagtctgg cctgtaccgt cagcgccatt gacgctgcgc
P0121603367  .......... .......... .......... .......... ..........
P0121604367  .......... .......... .......... .......... ..........
P0121614367  ACCAGACAAT AATAGTCTGG CCTGTACC.. .......... ..........
P1231409367  CCAGACATAT GATCTGG... .......... .......... ..........
P1390516367  ACCAGACAAT AATAGTCTGG CC........ .......... ..........
P1390520367  ACCAGACAAT AATAGTCTGG CC........ .......... ..........
P1390501367  ACCAGACAAT AATAGTCTGG CC........ .......... ..........
P1390503367  ACCAGACAAT AATAGTCTGG CC........ .......... ..........
P1390608367  ACCAGACAAT AATAGTCTGG CCTGTACC.. .......... ..........
P1390610367  ACCAGACAAT AATAGTCTGG CCTGTACCGT CA........ ..........
P1390611367  ACCAGACAAT AATAGTCTGG CCTGTACCGT A......... ..........
```

```
P1396b09367  ACCAGACAAT AATAGTCTGG CCTGTACCGT CAGCGCCATT GACGCTGCGC
P1396b12367  accagacaat aatag..... .......... .......... ..........

             251                                 284
P1081827367  CC........ .......... .......... ....
P1081828367  .......... .......... .......... ....
P1081107367  .......... .......... .......... ....
P1081501367  .......... .......... .......... ....
P1081504367  .......... .......... .......... ....
P1081906367  .......... .......... .......... ....
P0120810367  ccatagtgct tcctgctgct cccaagaacc aagg
P0121603367  .......... .......... .......... ....
P0121604367  .......... .......... .......... ....
P0121614367  .......... .......... .......... ....
P1231409367  .......... .......... .......... ....
P1390516367  .......... .......... .......... ....
P1390520367  .......... .......... .......... ....
P1390501367  .......... .......... .......... ....
P1390503367  .......... .......... .......... ....
P1390608367  .......... .......... .......... ....
P1390610367  .......... .......... .......... ....
P1390611367  .......... .......... .......... ....
P1396b09367  CCATAGTGCT TC........ .......... ....
P1396b12367  .......... .......... .......... ....
```

Acknowledgements

I would like to thank my supervisor Andrew Leigh Brown.

My thanks go to those at the Centre for HIV Research, Edinburgh University, especially to Pamela Robertson, Sandy Clelland, and my collaborators in this project - Robert Walker and Elizabeth Harvey. Also many thanks to Chris Wade who took upon himself the difficult task of setting-up the HIV computer network. I thank Marilyn Moore for the use of 'Cricket Graph', and thank Alicia Alonzo, who has provided constructive criticism of drafts of this thesis.

I could not have managed without the support from Petra zur Lage, Richard Waites, Jane Carlton and Bruce Wedgwood-Oppenheim. And from Catriona Baker for many pints in the Union bar.

However, my particular thanks are to the staff and patients at Malham House Day Hospital, Leeds (April 1995-January 1996), without whom I could not have finished this thesis.

References

Adachi, A., Gendelman, H., Koenig, S., Folks, T., et al (1986). Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. J Virol 59 284-291.

Aiken, C., Konner, J., Landau, N., Lenburg, M., Trono, D. (1994). Nef induces CD4 endocytosis: requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. Cell 76 853-864.

Ait-Khaled, M., and Emery, V. (1993). Sequence variation within the human immunodeficiency virus V3 loop at seroconversion. J Med Virol 41 270-274.

Albert, J., Abrahamson, B., Nagy, K., et al. (1990). Rapid development of isolate specific neutralising antibodies after primary HIV-1 infection and consequent emergence of virus variants which resist neutralisation by autologous sera. AIDS 4 107-112.

Alizon, J., Wain-Hobson, S., Montagnier, L., et al. (1986). Genetic Variability of the AIDS virus: nucleotide sequence analysis of two isolates from african patients. Cell 46 63-74.

Alting-Mees, M., and Short, J. (1989). pBluescript II: gene mapping vectors. Nucleic Acids Research 17 9494.

Andeweg, A., Leeflang, P., Osterhaus, A., Bosch, M. (1993). Both V2 and V3 regions of the human immunodeficiency virus type 1 surface glycoprotein functionally interact with other envelope regions in syncytium formation. J Virol 67 3232-3239.

Arendrup, M., Nielson, C., Hansen, J., Pedersen, C., Mathiesen, L., Nielson, J. (1992) Autologous HIV-1 neutralizing antibodies: emergence of neutralization-resistant escape virus and subsequent development of escape virus neutralizing antibodies. J. Acqir Immune Defic Syndr 5 303-307.

Arendrup, M., Sonnerborg, A., Svennerholm, B., et al. (1993). Neutralizing antibody response during human immunodeficiency virus type 1 infection: type and group specificity and viral escape. J Gen Virol 74 855-863.

Ariyosha, K., Harwood, E., Cheingsong-Popov, R., Weber, J. (1992). Is clearance of HIV-1 viraemia at seroconversion mediated by neutralising antibodies ?. Lancet 340 1257-1259.

Arthur, L., Bess, J., Sowder, R., et al. (1993). HLA DR and β2 microglobulin are bound to human and simian immunodeficiency viruses and are involved in infection. Science 258 1935-1938.

Asjo, B., Morfeldt-Manson, L., Albert, J., et al. (1986). Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection Lancet ii 660-662.

Baba, T., Jeong, Y., Pennick, D., et al. (1995). Pathogenicity of live, attenuated SIV

after mucosal infection of neonatal macaques. Science 267 1820-1825.

Back, N., Smit, L., Schutten, M., et al. (1993). Mutations in human immunodefiency virus type 1 gp41 affect sensitivity to neutralization by gp120 antibodies. J Virol 67 6897-6902.

Baier, M., Dittmar, M., Cichutek, K. and Kurth, R. (1991). Development in vivo of genetic variability of simian immunodeficiency virus. PNAS USA 88 8126-8130.

Balfe, P., Simmonds, P., Ludlam, et al. (1990). Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: rate of sequence change and low frequency of inactivating mutations. J Virol 64 6221-6233.

Ball, J., Holmes, E., Whitwell, H., Desselberger, U. (1994). Genomic variation of human immunodeficiency virus type 1 (HIV-1): molecular analyses of HIV-1 in sequential blood samples and various organs obtained at autopsy. J Gen Virol 75 876-878.

Barbas, C., Pyati, J., Koduri, R., et al. (1994). Primary isolates of human immunodeficiency virus type 1 are efficiently neutralized by a recombinant human monoclonal antibody. Science 266 1024-1027.

Barin, F., McLane, M., Allen, J., et al (1985). Virus envelope protein of HTLV-III represents a major target antigen for antibodies in AIDS patients. Science 228 1094-1096.

Bell, J., Busuttil, A., Ironside, J., et al. (1993). Human immunodeficiency virus and the brain: investigation of virus load and neuropathological changes in pre-AIDS subjects. J Infec Dis 168 818-824.

Benjamin, D., Berzofsky, J., East, I., et al. (1984). The antigenic structure of proteins: a reappraisal. Ann Rev Immunol 1984 2 67-101.

Benko, D., Schwartz, S., Pavlakis, G., Felber, B. (1990). A novel human immunodeficiency virus type 1 protein, tev, shares sequences with tat, env, and rev proteins. J Virol 64 2505-2518.

Bergeron, L., Sullivan, N., Sodroski, J. (1992). Target cell specific determinants of membrane fusion within the human immunodeficiency virus type 1 gp120 third variable region and gp41 amino terminus. J Virol 66 2389-2397.

Bergseid, M., et al. (1991) Stratagies 4 34-35

Berkower, I., Smith, G., Giri, C., and Murphy, D. (1989). Human immunodeficiency virus 1. Predominance of a group specific neutralising epitope that persists despite genetic variation. J Exp Med 170 1681-1695.

Bieniasz, P. D., Ariyoshi, K., Bourelly, M., Bloor, S. et al. (1993). Variable relationship between proviral DNA load and infectious virus titre in the peripheral

blood mononuclear cellsof HIV-1-infected individuals. AIDS 7 803-806.

Bloom, B. (1979). Games parasites play: how parasites evade immune surveillance. Nature 279 21-26.

Bolognesi, D. (1993). Human immunodeficiency virus vaccines. Adv Virus Res 42 103.

Bonhoeffer, S., Holmes, E., and Nowak, M. (1995). Causes of HIV Diversity. Nature 376 125.

Botarelli, P., Houlden, B., Haigwood, N., et al. (1991). N-glycosylation of HIV-gp120 may constrain recognition by T lymphocytes. J Immunol 147 3128-3132.

Bou-Habib, D., Roderiquez, G., Orevecz, T., Berman, P., et al. (1994). Cryptic nature of envelope V3 region epitopes protects primary human immunodeficiency virus type 1 from antibody neutralization. J Virol 68 6006-6013.

Bour, S., Schubert, U., and Strebel, K. (1995). The human immunodeficiency virus type 1 Vpu protein specifically binds to the cytoplasmic domain of CD4: implication for the mechanism of degredation. J Virol 69 1510-1520.

Boyd, M., Simpson G., Cann., Johnson, M., Weiss, R. (1993). A single amino acid substitution in the V1-loop of human immunodeficiency virus type 1 gp120 alters cellular tropism. J Virol 67 3649-3652.

Braathen, R., Ramirez, G., Kunze, R., Gelderblom, H. (1987). Langerhans cells as primary target cells for HIV infection. Lancet 1094.

Brinchmann, J., Albert, J., Vartdal, F. (1991). Few infected CD4+ T cells but a high proportion of replication competent provirus copies in asymptomatic human immunodeficiency virus type 1 infection. J Virol 65 2019-2023.

Bukrinsky, M., Haggerty, M., Dempsey, M., et al. (1993). A nuclear localisation signal within HIV-1 matrix protein that governs infection of non-dividing cells. Nature 365 666-669.

Bullock, W., Fernandez, J. and Short, J. (1987). XL1-Blue: A high efficiency plasmid transforming recA E.coli strain with Beta-galactosidase selection. Biotechniques5 4 376-379.

Burns, D., and Desrosiers, R. (1991). Selection of genetic variants of simian immunodeficiency virus in persistently infected rhesus monkeys. J Virol 65 1843-1854.

Burns, D., Collignon, C., Desrosiers, R. (1993). Simian immunodeficiency virus mutants resistant to serum neutralization arise during persistent infection of rhesus monkeys J. Virol 67 4104-4113.

Cann, A., Churcher, J., Boyd, M., et al. (1992). The region of the envelope gene of human immunodeficiency virus type 1 responsible for determination of cell tropism. J Virol 66 305-309.

Cao, Y., Qin, L., Zhang, L., Safrit, J., and Ho, D. (1995). Virological and immunological characterization of long-term survivors of human immunodeficiency virus type 1. N Engl J Med 302 201-208.

Carpenter, S., Evans, L., Sevian, M., Chesebro, B. (1987). Role of the host immune response in selection of equine infectious anemia virus variants. J Virol 61 3783-3789.

Centers for Infectious Diseases (CDC): Classification System for human T-lymphotropic virus type III/lymphadenopathy-associated virus infections. (1986). MMWR 35 334-339.

CDC: Revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. (1993a). MMWR 41 (RR-17) 1-19.

CDC Update: Mortality attributable to HIV-1 infection among persons aged 25-44 years - United States, 1991 and 1992. (1993b) MMWR 42 869-872.

Chakrabarti, S., Robert-Guroff, M., Wong-Staal, et al. (1986). Expression of the HTLV-III envelope gene by a recombinant vaccinia virus. Nature 320 535-537.

Chalker, A., Leach, D., and Lloyd (1988). Escherichia coli sbcC mutants permit stable propagation of DNA replicons containing a long palindrome. *Gene* 71 201-205.

Chamat, S., Nara, P., Berquist, L., et al. (1992). Two major groups of neutralising antibodies exist in HIV-infected individuals: evidence for epitope diversity around the CD4 attachment site. J Immunol 149 649-654.

Chang, D., Sharp, P. (1989). Regulation by HIV Rev depends upon recognition of splice sites. Cell 59 789-795.

Charneau, P., Borman, A., Quillent, C. et al. (1994). Isolation and envelope sequence of a highly divergent HIV-1 isolate: definition of a new HIV-1 group. Virology 205 247-253.

Cheng-Mayer, C., Seto, D., Tateno, M., and Levy, J. A. (1988). Biological features of HIV-1 that correlate with virulence in the host. Science 240 80-82.

Cheng-Mayer, C., Weiss, C., Seto, D., Levy, J. (1989). Isolates of human immunodeficiency virus type 1 from the brain may constitute a special group of the AIDS virus. PNAS USA 86 8575-8579.

Cheng-Mayer, C., Quiroga, M., Tung, et al. (1990). Viral determinants of human immunodeficiency virus type 1 T cell or macrophage tropism, cytopathogenicity, and

CD4 antigen modulation. J Virol 4390-4398.

Cheng-Mayer, C., Shioda, T, and Levy, J. (1991). Host range, replicative, and cytopathic properties of human immunodeficiency virus type 1 are determined by very few amino acid changes in tat and gp120. J Virol 65 6931-6941.

Chesebro, B., Nishio, J., Perryman, et al. (1991). Identification of human immunodeficiency virus envelope gene sequences influencing viral entry into CD4-positive HeLa cells, T-leukemia cells, and macrophages. J Virol 64 5782-5789.

Chesebro, B., Wehrly, K., Nishio, J., Perryman, S. (1992). Macrophage tropic human immunodeficiency isolates from different patients exhibit unusual V3 envelope sequence homology in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. J Virol 66 6547-6554.

Cichutek, K., Norley, S., Linde, R., et al. (1991). Lack of HIV-1 V3 region sequence diversity in two haemophiliac patients infected with a putative biological clone of HIV-1. AIDS 5 1185-1187.

Cichutek, K., Merget, H., Norley, et al. (1992). Development of a quasispecies of human immunodeficiency virus type 1 in vivo. PNAS USA 89 7365-7369.

Clark, S., Saag, M., Decker, W., et al. (1991). High titers of cytopathic virus in plasma of patients with primary HIV-1 infection. N Engl J Med 324 954-960.

Clavel, F., Hoggan, M., Willey, R., et al. (1989). Genetic Recombination of human immunodeficiency virus. J Virol 63 1455-1459.

Clerici, M. (1993). Cell-mediated immunity in HIV infection. AIDS S135-140.

Cocchi, F., DeVico, A., Garzino-Demo, A., et al. (1995). Indentification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells. Science 270 1811-5.

Collin, M., Illei, P., James, W., Gordon, S. (1994). Definition of the range and distribution of human immunodeficiency virus macrophage tropism using PCR-based infectivity measurements. J Gen Virol 75 1597-1603.

Collman, R., Balliet, J., Gregory, S., et al. (1992). An infectious molecular clone of an unusual macrophage tropic and highly cytopathic strain of human immunodeficiency virus type 1. J Virol 66 7517-7521.

Connor, R., Notermans, D., Mohri, H., Cao, Y., Ho, D. (1993a). Biological cloning of functionally diverse quasispecies of HIV-1. AIDS Res Hum Retroviruses 9 541-546.

Connor, R., Mohri, H., Cao, Y., Ho, D. (1993b). Increased viral burden and cytopathicity correlate temporally with CD4+ T-lymphocyte decline and clinical progression in human immunodeficiency virus type 1-infected individuals. J Virol 67

1772-1777.

Connor, R., and Ho, D. (1994). Human immunodeficiency virus type 1 variants with increased replication capacity develop during the asymptomatic stage before disease progression. J Virol 68 4400-4408.

Cornelissen, M., Mulder-Kampinga, G., Veenstra, J., et al. (1995). Syncytium-inducing (SI) phenotype suppression at seroconversion after intramuscular inoculation of a non-syncytium-inducing/SI phenotypically mixed human immunodeficiency virus population. J Virol 69 1810-1818.

Cullen, B. (1991). Human immunodeficiency virus as a prototype complex retrovirus. J Virol 65 1053-1056.

Daar, E., Li, X., Moudgil, T., Ho, D. (1990). High concentrations of recombinant soluble CD4 are required to neutralize primary human immunodeficiency type 1 isolates. PNAS USA 87 6574-6578.

Daar, E., Moudgil, T., Meyer, R., Ho, D. (1991). Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection. N Engl J Med 324 961-964

Dalgleish, A., Beverley, P., Clapham, P, et al. (1984). The CD4(T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature 312 763-765.

Daniel, M., Kirchoff, F., Czajak, S., et al. (1992). Protective effects of a live attenuated SIV vaccine with a deletion in the nef gene. Science 258 1938- 1941.

Dayton, E., Konings, D., Lim, S., et al. (1993). The RRE of the human immunodeficiency virus type 1 contributes to cell-type-specific viral tropism. J Virol 67 2871-2878.

De Jong, J-J, Goudsmit, J., Keulen, W., et al. (1993a). Human immunodeficiency virus type 1 clones chimeric for the envelope V3 domain differ in syncytium formation and replication capacity. J Virol 66 757-765.

De Jong, J. J., de Ronde, A., Keulen, W., et al. (1993b). Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. J Virol 66 6777-6780.

Delassus S., Cheynier, R., Wain-Hobson, S. (1992). Nonhomogeneous distribution of human immunodeficiency virus type 1 provirus in the spleen. J Virol 66 5642-5645.

Delwart, E., Shpaer, E., Louwagie, J., et al. (1993). Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 env genes. Science 262 1257-1261.

Delwart, E., Sheppard, H., Walker, B., et al. (1994). Human immunodeficiency virus type 1 evolution in vivo tracked by DNA heteroduplex mobility assays. J Virol 68

6672-6683.

Deng, H., Liu, R., Ellmeier, W., et al. (1996). Identification of a major co-receptor for primary isolates of HIV-1. Nature 381 661-666.

Desrosiers, R. (1992). HIV with multiple gene deletions as a live attenuated vaccine for AIDS. AIDS Res Hum Ret 8 411-421.

Desrosiers, R. (1994). Yes, it's time to consider the use of a live-attenuated virus vaccine against HIV-1. J NIH Res 6 54-59.

Devereux, J., Haeberil, P., Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucleic Acid Research 12 387-395.

Diaz, R., Sabino, E., Mayer, A., et al. (1995). Dual human immunodeficiency virus type 1 infection and recombination in a dually exposed transfusion recipient. J Virol 69 3273-3281.

Di Marzo Varonese, F., Reitz, M., Gupta, G., et al. (1993). Loss of neutralizing epitope by a spontaneous point mutation in the V3 loop of HIV-1 from an infected laboratory worker. J Biol Chem 268 25894-25901.

Domingo, E., Sabo, D., Tanaguchi, T., Weissmann, C. (1978). Nucleotide sequence heterogeneity of an RNA phage population. Cell 13 735.

Domingo, E., Diez, J., Martinez, et al. (1993). New observations on antigenic diversification of RNA viruses. Antigenic variation is not dependent on immune selection. J Gen Virol 74 2039-2045.

Donaldson, Y., Bell, J., Brettle, R., et al. (1994a). Redistribution of HIV outside the lymphoid system with onset of AIDS. Lancet 343 382-385.

Donaldson, Y., Bell, Y., Holmes, E., et al. (1994b). In Vivo distribution and cytopathology of variants of human immunodeficiency virus type 1 showing restricted sequence variability in the V3 loop. J Virol 68 5991-6005.

Dragic, T., Litwin, V., Allaway, G., et al. (1996). HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. Nature 381 667-673.

Du, B., Wolf, A., Lees, S., Terwilliger, E. (1993). Changes in the host range and growth potential of an HIV-1 clone are conferred by the vpu gene. Virology 195 260-264.

Dunning, A. M., Talmud, P., and Humphries, S. E. (1988). Errors in the polymerase chain reaction. Nucleic Acid Research 16 10393.

Eigen, M., McCaskill, J., Schuster, P. (1988). Molecular quasi-species. J Phys Chem 92 68881.

Eigen, M., Nieselt-Struwe, K. (1990). How old is the immunodeficiency virus?. AIDS 4 (suppl 1) S85-S93.

Embretson, J., Zupanici, M., Ribas, J., et al. (1993). Massive covert infection of helper T lymphocytes and macrophages by HIV during the incubation period of AIDS. Nature 362 359-362.

Emini, E., Schlief, W., Nunberg, J., et al. (1991). Prevention of HIV-1 infection in chimpanzees by gp120 V3 domain-specific monoclonal antibody. Nature 355 728-730.

Ennis, P. D., Zemmur, R. D., Salter, R. D., and Parham, P. (1990). Rapid cloning of HLA-, B cDNA by using the polymerase chain reaction: frequency and nature of errors produced in amplification. PNAS USA 87 2833-2837.

Epstein, L., Kuiken, C., Blumberg, B., Hartman, S., et al. (1991). HIV-1 V3 domain variation in brain and spleen of children with AIDS: tissue-specific evolution within host-determined quasispecies. Virology 180 583-590.

Felber, B, and Pavlakis, G. (1993). Molecular biology of HIV-1: positive and negative regulatory elements important for virus expression. AIDS 7 S51-62.

Feng, Y., Broder, C., Kennedy, P. and Berger, E. (1996). HIV-1 entry cofactor:Functional cDNA cloning of a seven-transmembrane G protein-coupled receptor. Science 272 872-877.

Fenner, F., Kerr, P. (1994). Evolution of poxviruses, including the coevolution of virus and host in myxomatosis. In *The evolutionary biology of viruses*, ed S Morse 273-292 New York: Raven Press.

Fenyo, EM., Morfeldt-Manson, L., Chiodi, F., et al. (1988). Distinctive replicative and cytopathic characteristics of human immunodeficiency isolates. J Virol 62 4414-4419.

Fenyo, EM., Albert, J., Asjo, B. (1989). Replicative capacity, cytopathic effect and cell tropism of HIV. AIDS 3 (suppl 1) S5-S12.

Fenyo, EM. (1994). Antigenic variation of primate lentiviruses in humans and experimentally infected macaques. Immunological Reviews 140 131-143.

Fiore, J., Bjorndal, A., Peipke, K., Stefano, M., et al. (1994). The biological phenotype of HIV-1 is usually retained during and after sexual transmission. Virology 204 297-303.

Fisher, A., Collalti, E., Ratner, L., Gallo, R. C., and Wong-Staal, F. (1985). A molecular clone of HTLV-III with biological activity. Nature 316 262-265.

Fisher, A., Feinberg, M., Josephs, S., Harper, M., et al. (1986). The trans-activator gene of HTLV-III is essential for virus replication. Nature 320 367-371.

Fitch, W., Leiter, J., Li, X., and Palese, P. (1991). Positive Darwinian evolution in human influenza A viruses. PNAS USA 88 4270-4274.

Fouchier, R. A. M., Groenink, M., Koostra, N., et al. (1992). Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. J Virol 66 3138-3187.

Freed, E., Myers, D., Risser, R. (1990). Characterization of the fusion domain of the human immunodeficiency virus type 1 envelope glycoprotein gp41. PNAS USA 87 4650-4654.

Freed, E., Martin, M. (1996). Domains of the human immunodeficiency virus type 1 matrix protein and gp41 cytoplasmic tail required for envelope incorporation into virions. J Virol 70 341-351.

Fujita, K., Silver, P., Peden, K. (1992). Changes in both gp120 and gp41 can account for increased growth potential and expanded host range of human immunodeficiency virus type 1. J Virol 66 4445-4451.

Fultz, P., McClure, H., Anderson, D., Switzer, W. (1989). Identification and biological characterization of an acutely lethal variant of simian immunodeficiency virus from sooty mangabeys (SIV/SMM). AIDS Res Hum Ret 5 397-409.

Fung, M. S., Sun, C.R., Gordon, W.L., et al. (1992). Identification and characterisation of the neutralisation site within the second variable region of human immunodeficiency virus type 1 gp120. J Virol 66 848-856.

Gallaher, W. (1987). Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus. Cell 50 327-328.

Gallaher, W., Ball, J., Garry, R., et al. (1989). A general model for the transmembrane proteins of HIV and other retroviruses. AIDS Res Hum Ret 5 431-440.

Gao, F., Yue, L., Robertson, D., et al. (1994). Genetic diversity of human immunodeficiency virus type 2: evidence for distinct sequence subtypes with differences in virus biology. J Virol 68 7433-7447.

Gao, F., Morrison, S., Robertson, D., et al. (1996). Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type sequence subtypes A through G. J Virol 70 1651-1667.

Garcia, J., Wu, F., Mitsuyasu, R., Gaynor, R. (1987). Interaction of cellular proteins involved in the transcriptional regulation of the human imunodeficiency virus. EMBO Journal 6 3761-3770.

Gartner, S., Popovic, M. (1990). Macrophage tropism of HIV-1. AIDS Res Hum Ret 6 1017-1021.

Gazdar, A., Carney, D., Bunn, P., et al. (1980). Mitogen requirements for the in vitro propagation of cutaneous T-cell lymphomas. Blood 55 409-419.

Gelderblom, H. (1991). Assembly and morphology of HIV: potential effect of structure on viral function AIDS 5 617-638.

Gendelman, H., Baca, L., Husayni, H., et al. (1990). Macrophage-HIV interaction: viral isolation and target cell tropism. AIDS 4 221-228.

Geraghty, R., Panganiban, A. (1993). Human immunodeficiency virus type 1 vpu has a CD4- and an envelope glycoprotein-independent function. J Virol 67 4190-4194.

Gibbs, J., Regier, D., Desrosiers, R. (1994). Construction and in vitro properties of HIV-1 mutants with deletions in 'nonessential' genes. AIDS Res Hum Retroviruses 10 343-350.

Gluzman, Y. (1981). SV40-transformed simian cells support the replication of early SV40 mutants. Cell 23 175-182

Goodenow, M., Huet, T., Saurin, W., Kwok, S., et al. (1989). HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. J Acq Imm Def Syn 2 344-352.

Gorny, M., Moore, J., Conley, A., et al. (1994). Human anti-V2 monoclonal antibody that neutralizes primary but not laboratory isolates of human immunodeficiency virus type 1. J Virol 68 8312-9320.

Goudsmit, J., Debouck, C., Meloen R., et al. (1988). Human immunodeficiency virus type 1 neutralization epitope with conserved architecture elicits early type-specific antibodies in experimentally infected chimpanzees. PNAS 85 4478-4482.

Groenink, M., Fouchier, R., de Goede, R., et al. (1991). Phenotypic heterogeneity in a panel of infectious molecular human immunodeficiency type 1 clones derived from a single individual. J Virol 65 1968-1975.

Groenink, M., Andeweg, A., Fouchier, R., et al. (1992). Phenotype associated env gene variation among eight related human immunodeficiency type 1 clones: evidence for in vivo recombination and determinants of cytotropism outside the V3 domain. J Virol 66 6175-6180.

Groenink, M., Fouchier, R., Broersen, S., Baker, C., et al. (1993). Relation of phenotype evolution of HIV-1 to envelope V2 configuration. Science 260 1513-1516.

Gruters, R., Otto, S., Al, S., et al. (1991). Non-mitogenic T-cell activation signals are sufficient for induction of HIV transcription. Eur J Immunol 21 167-172.

Haase, A. (1986). Pathogenesis of lentiviral infection. Nature 322 130-136.

Hahn, B., Shaw, G., Taylor, M., et al. (1986). Genetic variation in HTLV-III/LAV

over time in patients with AIDS or at risk for AIDS. Science 232 1548-1553.

Haigwood, N., Mason, O., York-Higgins, D., et al. (1993). HIV-1 envelope sequence diversity in long-term asymptomatic patients. Vaccines 93 233-242

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. j Mol Biol 166 557-580.

Hattori, N., Michaels, F., Fargnoli, K., et al. (1990). The human immunodeficiency virus type 2 vpr gene is essential for productive infection of human macrophages. PNAS USA 87 8080-8084.

Heaphy, S., Dingwall, C., Ernberg, I., Gait, M., et al. (1990). HIV-1 regulator of virion expression (Rev) protein binds to an RNA stem-loop structure located wthin the Rev response element region. Cell 60 685-693.

Heinzinger, N., Bukrinsky, M., Haggarty, S., et al. (1994). The vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acid in non-dividing host cells. PNAS USA 91 7311-7315.

Heinzinger, N., Bukrinsky, M., and Haggerty, J. (1994). The Vpr protein of HIV-1 influences nuclear localisation of viral nucleic acids in non-dividing host cells. PNAS USA 91 7311-7315.

Helseth, E., Kowalski, M., Gabuzda, D., et al. (1990a). Rapid complementation assay measuring replicative potential of human immunodeficiency virus type 1 envelope glycoprotein mutants. J Virol 64 2416-2420.

Helseth, E., Olshevsky, U., Gabuzda, D., et al. (1990b). Changes in the transmembrane region of the human immunodeficiency virus type 1 gp41 envelope glycoprotein affect membrane fusion. J Virol 64 6314-6318.

Hirsch, I., Spire, B., Tsunetsugu-Yokota, Y., et al. (1990). Differences in replication and cytopathogenicity of human immunodefiency virus type 1 (HIV-1) are not determined by long terminal repeats (LTR). Virology 177 759-763.

Ho, D., Kaplan, J., Rackauskas, I., et al. (1988). Second conserved domain of gp120 is important for HIV infectivity and antibody neutralisation. Science 239 1021-1023.

Ho, D., Moudgil, T., Alam, M. (1989). Quantitation of human immunodeficiency virus type 1 in the blood of infected persons. N Engl J Med 321 1621-1625.

Hogland, S., Ohagen, A., Lawrence, K., Clarkson, K., Gabuzda, D. (1994). Role of vif during packaging of the core of HIV-1. Virology 201 349-355.

Holmes, D. and Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. Anal Biochem 114 193-197.

Holmes, E., Zhang, L.Q., Simmonds, P., et al. (1992) Convergent and divergent

sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. PNAS USA 89 4835-4839.

Holmes, E., Zhang, LQ, Robertson, P., et al. (1995). The molecular epidemiology of HIV-1 in Edinburgh Scotland. J Infect Dis 171 45-53.

Homsy, J., Meyer, M., Tateno, M., Clerkson, S., Levey; J. (1989). The Fc and not CD4 receptor mediates enhancement of HIV infection in human cells. Science 244 1357-1360.

Homsy, J., Meyer, M., Levey, J. (1990). Serum enhancement of human immunodeficiency virus (HIV) infection correlates with disease in HIV-infecetd individuals. J Virol 64 1437-1440.

Howell, R., Fitzgibbon, J., Noe, M., et al. (1991). In vivo sequence variation of the human immunodeficiency virus type 1 env gene: evidence for recombination among variants found in a single individual. AIDS Res Hum Retroviruses 7 869-876.

Hu, SL., Kosowski, S., Dalrymple, J. (1986). Expression of AIDS virus envelope gene in recombinant vaccinia viruses. Nature 320 537-540.

Hu, W., Temin, H. (1990). Retroviral recombination and reverse transcription. Science 250 1227-1233.

Hwang, S., Boyle, H., Lyerly, H., and Cullen, B. (1991). Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. Science 253 71-74.

Itescu, S., Simonelli, P., Winchester, R., and Ginsberg, H. (1994). Human immunodeficiency virus type 1 strains in the lungs of infected individuals evolve independently from those in peripheral blood and are highly conserved in the C-terminal region of the envelope V3 loop. PNAS USA 91 11378-11383.

Iverson, A., Shpaer, E., Rodrigo, A., et al. (1995). Persistance of attenuated rev genes in a human immunodeficiency virus type 1-infected asymptomatic individual. J Virol 69 5743-5753.

Ivey-Hoyle, M., Culp, J., Chaikin, M., et al. (1991). Envelope glycoprotein from biologically diverse isolates of immunodeficiency viruses have widely different affinities for CD4. PNAS USA 88 512-516.

Javaherian, K., Langlois, A. J., McDanal, C., et al. (1989). Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. PNAS 86 6768-6772.

Javaherian, K., Langlois, A., LaRosa, G., et al. (1990). Broadly neutralising antibodies elicited by the hypervariable neutralising determinant of HIV-1. Science 250 1590-1593.

Johnson, P., Hamm, T., Goldstein, S., Kitov, S., Hirsch, V. (1991). The genetic fate of molecularly cloned simian immunodeficiency virus in experimentally infected macaques. Virology 185 217-228

Jones, K., Kadonaga, J., Luviw, P., Tijan, R. (1986). Activation of the AIDS retrovirus promoter by the cellular transcription factor Sp1. Science 232 755-758

Joyce, C. and Grindley. (1984). Method for determining whether a gene of Escherichia coli is essential: application to the polA gene. J Bacteriol 158 636643.

Kang, C., Nara, P., Chamat, S., et al. (1991). Evidence for non-V3-specific neutralizing antibodies that interfere with gp120/CD4 binding in human immunodeficiency virus type 1-infected humans. PNAS USA 88 6171-6175.

Kannagi, M., Masuda, T., Hattori, T., et al. (1990). Interference with human immunodeficiency virus (HIV) replication by CD8+ T cells in peripheral blood leukocytes of asymptomatic HIV carriers. J Virol 64 3399-3406.

Kawamura, M., Ishiazaki, T., Ishimoto, A., Shioda, T., et al. (1994). Growth ability of human immunodeficiency virus type 1 auxiliary gene mutants in primary blood macrophage cultures. J Gen Virol 75 2427-2431.

Kayman, S., Wu, Z., Revesz, K., et al. (1994). Presentation of native epitopes in the V1/V2 and V3 regions of human immunodeficiency virus type 1 gp120 by fusion glycoproteins containing isolated gp120 domains. J Virol 68 400-410.

Kestler, H., Ringler, D., Mori, K., Panicali, D., et al (1991). Importance of the nef gene for maintenance of high virus loads and for the development of AIDS. Cell 65 651-662.

Kin, S., Ikeuchi, K., Byrn, R., Groopman, J., Baltimore, D. (1989). Lack of negative influence on viral growth by the nef gene of human immunodeficiency virus type 1. PNAS USA 86 9544-9548.

Klasse, P., McKeating, J., Schuten, M., Reitz, M., Robert-Guroff, M. (1993). An immune-selected point mutation in the transmembrane protein of human immunodeficiency virus type 1 (HXB2-env:Ala 581-(Thr)) decreases viral neutralization by monoclonal antibodies to the CD4-binding site. J Virol 67 332-337.

Klatzmann, D., Barre-Sinoussi, F., Nugeyre, T., et al. (1984a). Selective tropism of lymphadenopathy virus (LAV) for helper-induced T lymphocytes. Science 225 59-63.

Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., et al. (1984b). T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature 312 767-768.

Kliks, S., Wara, D., Landers, D., and Levy, J. (1994). Features of HIV-1 that could influence maternal-child transmission. JAMA 272 467-474.

Klimkait, T., Strebel, K., Hoggan, M., Martin, M., Orenstein, J. (1990). The human

immunodeficiency virus type 1-specific protein vpu is required for efficient virus maturation and release. J Virol 64 621-629.

Knight, S. (1990). HIV-1 infection of dendritic cells. Int Rev Immunol 6 163-175.

Koenig, S., Gendelman, H., Orenstein, J., Dal Canto, M., et al. (1986). Detection of AIDS virus in macrophages in brain tissue from AIDS patients with encephalopathy. Science 233 1089-1093.

Koito, A., Harrowe, G., Levy, J., Cheng-Mayer, C. (1994). Functional role of the V1-V2 region of human immunodeficiency virus type 1 envelope glycoprotein in infection of primary macrophages and soluble CD4 neutralization. J Virol 68 2253-2259.

Koot, M., Vos, A., Keet, R., et al. (1992). HIV-1 biological phenotype in long-term infected individuals evaluated with an MT-2 cocultivation assay. AIDS 6 49-54.

Koot, M., Keet, I., Vos, A., et al. (1993). Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS. Ann Intern Med 118 681-688.

Korber, B., Kuntsman, K., Patterson, B., et al. (1994). Genetic differences between blood- and brain-derived viral sequences from HIV-1 infected patients: evidence of conserved elements in the V3 region of the envelope protein of brain-derived sequences. J Virol 68 7467-7481.

Kostrikis, G., Bagdades, E., and Cao, Y. (1995). Genetic analysis of HIV-1 strains from patients in Cyprus: identification of a new subtype designated subtype I. J Virol 69 6122-6130.

Koup, R., Safrit, J., Cao, Y., et al. (1994). Temporal association of cellular immune response with the initial control of viremia in primary HIV-1 syndrome. J Virol 68 4650-4655.

Kuiken, C., Zwart, G., Baan, E., Coutinho, R, van den Hoek, J., Goudsmit, J. (1993). Increasing antigenic diversity of the V3 variable domain of the human immunodeficiency virus envelope protein in the course of AIDS. PNAS USA 90 9061-9065.

Kunkel, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. PNAS USA 82 488-492.

Kusumi, K., Conway, B., Cunningham, S., et al. (1992). Human immunodeficiency virus type 1 envelope gene structure and diversity in vivo and after co-cultivation in vitro. J Virol 66 875-885.

Laemmli, U. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227 660-685.

Langhoff, E., Terwilliger, E., Bos, H., Kalland, K., et al. (1991).Replication of human immunodeficiency virus type 1 in primary dendritic cells PNAS USA 88 7998-8002.

Larder, B., Darby, G., Richmann, D. (1989). HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. Science 243 1731-1734.

LaRosa, G., Davide, J., Weinhold, K., et al. (1990). Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. Science 249 932-935.

Lasky, L., Nakamura, G., Smith, D., Fennie, C., et al. (1987). Delineation of a region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with the CD4 receptor. Cell 50 975-985.

Laspia, M., Rice, A., Mathews, M. (1989). HIV-1 tat protein increases initiation and stabilizes elongation. Cell 59 283-292.

Laver, W., Air, G., Webster, R. and Smith-Gill. (1990). Epitopes on protein antigens: misconceptions and realities. Cell 61 553-556.

Leigh Brown, A. (1991). Sequence variability in immunodeficiency viruses: pattern and process in viral evolution. AIDS (Suppl 2) 5 S35-S42.

Leigh Brown, A. (1996). Analysis of selection in HIV-1 env gene sequences reveals evidence for a low effective number in the viral population. Submitted.

Leigh Brown, A. and Holmes, E. (1994).Evolutionary biology of human immunodeficiency virus. Ann Rev Ecol Syst 25 127-165.

Leigh Brown, A., and Cleland, A.(1996). Independent evolution of the env and pol genes of HIV-1 during zidovudine therapy. AIDS 10 1067-1073.

Lenburg, M., Landau, N. (1993). Vpu induced degradation of CD4: requirement for specific amino acid residues in the cytoplasmic domain of CD4. J Virol 67 7238-7245.

Leonard, C., Spellman, M., Riddle, L., et al. (1990). Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. J Biol Chem 265 10373-10382.

Levy, J., Cheng-Mayer, C., Dina, D., Luciw, P. (1986). AIDS retrovirus (ARV-2) clone replicates in transfected human and animal fibroblasts. Science 232 998-1001.

Levy, J. (1993). Pathogenesis of human immunodeficiency virus infection. Microbiol. Rev 57 183.

Lewis, P., and Emerman, M. (1994). Passage through mitosis is required for oncoretroviruses but not for human immunodeficiency virus. J Virol 68 570-516.

Li, W., Tanimura, M., Sharp, P. (1988). Rates and dates of divergence between AIDS virus nucleotide sequences. Mol Biol Evol 5 313-330.

Li, Y., Hui, H., Price, R., Sharp, P., et al. (1992). Complete nucleotide sequence, genome organisation and biological properties of human immunodeficiency virus type 1 in vivo: evidence for limited defectiveness and complementation. J Virol 66 6587-6600.

Looney, D. J., Fisher A. J., Putney, S. D., et al. (1988). Type restricted neutralisation of molecular clones of HIV. Science 241 357-359.

Louwagie, J., McCutchan, F., Peeters, M., et al. (1993). Phylogenetic analysis of gag genes from 70 international isolates provides evidence for multiple genotypes. AIDS 7 769-780.

Lu, W., Shih, W., Tourani, JM., et al. (1993). Lack of isolate-specific neutralising activity is correlated with an increased viral burden in rapidly progressing HIV-1-infected patients. AIDS 7 S91-S99.

Ludlam, C. A., Tucker, J., Steel C. M., et al. (1985). Human T-lymphotropic virus type III (HTLV-III) in seronegative haemophiliacs after transfusion of factor VIII. Lancet ii 233-236.

Lukashov, V., Kuiken, C., Goudsmit, J. (1995). Intrahost human immunodeficiency virus type 1 evolution is related to length of the immunocompetant period. J Virol 69 6911-6916.

Lutley, R., Petursson, G., Palsson, P., et al.(1983). Antigenic drift in visna: virus variation during long-term infection of icelandic sheep. J Gen Virol 64 1433-1440.

Macatonia, S., Patterson, R., Pinching, A., Knight, S. (1990). Dendritic cell infection, depletion and dysfunction in HIV-infected individuals. Immunology 71 38-45.

Maddon, P., Daldleish, A., McDougal, J., Clapham, P., Weiss, R., Axel, R. (1984). The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and brain. Cell 47 333-348.

Malim, M., Hauber, J., Fenrick, R., Cullen, B. (1988). Immunodeficiency virus rev trans-activator modulates the expression of the viral regulatory genes. Nature 335 181-183.

Malim, M., Hauber, J., Le, SY., Maizel, J., Cullen, B. (1989). The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature 338 254-247.

Malim, M., Tiley, L., McCarn, D. et al (1990). HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. Cell 60 675-683.

Malim, M., and Cullen, B. (1991). HIV-1 structural gene expression requires the binding of multiple rev monomers to the viral RRE: implications for HIV-1 latency. Cell 65 241-248.

Mann, D., O'Brien, J., Gilbert, D., et al. (1989). Origin of the HIV-susceptible human CD4+ cell line H9. AIDS Res Hum Ret 5 253-255.

Mansky, L., and Temin, H. (1995). Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase J Virol 69 5087-5094.

Marciniak, R., Calnan, B., Frankel, A., Sharp, P. (1990). HIV-1 tat protein trans-activates transcription in vitro. Cell 63 791-802.

Mariani, R., and Skowronski, J. (1993). CD4 down-regulation by nef alleles from the human immunodeficiency virus type 1 infected individuals. PNAS USA 90 5549-5553.

Marlink, R., Kanki, P., Thior, I., Travers, K., et al. (1994). Reduced rate of disease development after HIV-2 infection as compared to HIV-1. Science 265 1587-1590.

Masuda, T., Matsushita, S., Kuroda, M., et al. (1990). Generation of neutralisation resistant HIV-1 in vitro due to amino acid interchanges of third hypervariable env region. J Immunol 145 3240-3246.

Mathews, T. (1994). Dilema of neutralization resistance to HIV-1 field isolates. AIDS Res Hum Ret 10 631-632.

Mathur, E., Shoemaker, D., Rombouts, J., Bergseid, M. and Nielson, K. (1992). Pfu DNA polymerase update. Strategies in molecular biology (Stratagene) 5 11-13.

McClelland, M., Hanish, J., Nelson, M., and Patel, Y. (1988). KGB: a single buffer for all restriction endonucleases. Nucleic Acids Research 16 364.

McClure, M., Marsh, M., Weiss, R. (1988). Human immunodeficiency virus infection of CD4-bearing cells occurs by a pH-independent mechanism. EMBO Journal 7 513-518.

McCune, J., Rabin, L., Feinberg, M., et al. (1988). Endoproteolytic cleavage of gp160 is required for activation of human immunodeficiency virus. Cell 53 55-67.

McKeating, J., Gow, J., Goudsmit, J., Pearl, L., Mulder, C., Weiss, R. (1989). Characterization of HIV-1 neutralization escape mutants. AIDS 3 777-784.

McKeating, J., McKnight, A., and Moore, J. (1991). Differential loss of envelope glycoprotein gp120 from virions of human immunodeficiency virus type 1 isolates: effects on infectivity and neutralisation. J Virol 65 852-860.

McKeating, J., Bennet, J., Zolla-Pazner, S., Schutten, M., Ashelford, S. et al (1993).

Resistance of a human serum-selected human immunodeficiency virus type 1 escape mutant to neutralization by CD4 binding site monoclonal antibodies is conferred by a single amino acid change in gp120. J Virol 67 5216-5225.

McKnight, A., Clapham, P., Goudsmit, J. et al. (1992). Development of HIV-1 group-specific neutralising antibodies after seroconversion. AIDS 6 799-802.

McNearney, T., Westervelt, P., Thielan, B., et al. (1990). Limited sequence heterogeneity among biologically distinct human immunodeficiency virus type 1 isolates from individuals involved in a clustered outbreak. PNAS USA 87 1971-1921.

McNearney, T., Hornickova, Z, Markham, R., et al. (1992). Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. PNAS USA 89 10247-10251.

Melton, D. A., Krieg, P. A., Rebagliati, M. R., et al. (1984). Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Research 12 7034-7057.

Meyerhans, A., Cheynier, R., Albert, J., et al. (1989). Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations. Cell 58 901-910

Meyerhans, A., Vartanian, J. P., and Wain Hobson, S. (1990). DNA recombination during PCR. Nucleic Acids Research 18 1687-1691.

Michael, N., Vahey, M., Burke, D., Redfield, R. (1992). Viral DNA and mRNA expression correlate with the stage of human immunodeficiency virus (HIV) type 1 infection in humans: evidence for viral replication in all stages of HIV disease. J Virol 66 310-316.

Miedema, F. (1992). Immunological abnormalities in the natural history of HIV infection: mechanisms and clinical relevance. Immunodef Rev 3 173-193.

Miller, M., Warmerdam, M., Gaston, I., et al. (1994). The human immunodeficiency virus type 1 nef gene product: a positive factor for viral infection and replication in primary lymphocytes and macrophages. J Exp Med 179 101-113.

Milich, L., Margolin, B., Swastrom, R. (1993). V3 loop of the human immunodeficiency virus type 1 Env protein: interpreting sequence variability. J Virol 67 5623-5634.

Mims, C. (1986). *The Pathogenesis of Infectious Disease*. Third Edition. Academic Press Ltd, 24-28 Oval Road, London NW1 7DX.

Modrow, S., Hahn, B., Shaw., Gallo, R. et al. (1987). Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions. J Virol 61 570-578.

Montefiori, D., Zhou, J., Barnes, B., et al. (1991). Homotypic antibody responses to fresh clinical isolates of human immunodeficiency virus. Virology 182 635-643.

Moore, J. (1990). Simple methods for monitoring HIV-1 and HIV-2 gp120 binding to soluble CD4 by enzyme-linked immunosorbant assay: HIV-2 has a 25-fold lower affinity than HIV-1 for soluble CD4. AIDS 4 297-305.

Moore, J. (1993). The reactivities of HIV-1+ human sera with solid-phase V3 loop peptides can be poor predictors of their reactivities with V3 loops on native gp120 molecules.

Moore, J., and Ho, D. (1995). HIV-1 neutralisation: the consequences of viral adaptation to growth on transformed T cells. AIDS 9 S117-S136.

Moore, J., Thali, M., Jameson, B., et al. (1993a). Immunochemical analysis of human immmunodeficiency virus type 1: probing the structure of the C4 and V4 domains and the interaction of the C4 domain with the V3 loop. J Virol 67 4785-4796.

Moore, J., Sattentau, Q., Yoshiyama, H., et al. (1993b). Probing the structure of the V2 domain of the human immunodeficiency virus type 1 surface glycoprotein gp120 with a panel of eight monoclonal antibodies: the human immune response to the V1 and V2 domain. J Virol 67 6136-6151.

Moore, J., Yoshiyama, H., Ho, D., et al. (1993c). Antigenic variation in gp120s from molecular clones of HIV-1 LAI. AIDS Res Hum Ret 9 1185-1193

Moore, J., Sattentau, Q, Wyatt, R., Sodroski, J (1994a). Probing the structure of the surface glycoprotein gp120 of human immunodeficiency virus type 1 with a panel of monoclonal antibodies. J Virol 68 469-484.

Moore, J., Cao, Y., Ho, D., and Koup, R. (1994b) Development of the anti-gp120 antibody response during seroconversion to human immunodeficiency virus type 1. J Virol 68 5142-5155.

Moore, J., McCutchan, F., Poon, SW., Mascola, J., et al. (1994c). Exploration of antigenic variation in gp120 from clades A through F of human immunodeficiency virus type 1 by using monoclonal antibodies. J Virol 68 8350-8364.

Moore, J., Cao, Y., Qing, L., et al. (1995). Primary isolates of human immunodeficiency virus type 1 are relatively resistant to neutralization by monoclonal antibodies to gp120 and their neutralization is not predicted by studies with monomeric gp120. J Virol 69 101-109.

Moore, J., Cao, Y., Leu, J., Qin, L., et al. (1996). Inter- and intraclade neutralisation of human immunodeficiency virus type 1: genetic clades do not correspond to neutralisation serotypes but partially correspond to gp120 antigenic serotypes. J Virol 70 427-444.

Mosier, D., Gulizia, R., Maclasaac, P., Torbett, B., and Levy, J. (1993). Rapid loss

of CD4+ T cells in human-PBL-SCID mice by noncytopathic HIV isolates. Science 260 689-692.

Mulder-Kampinga, G., Kuiken, C., Dekker, J., et al. (1993). Genomic human immunodeficiency virus type 1 RNA variation in mother and child following intra-uterine virus transmission. J Gen Virol 74 1747-1756.

Mullis, K. and Faloona, F. (1987). Specific synthesis of DNA via a polymerase-catalysed chain reaction. Meth in Enzymol 155 335-351

Muster, T, Guinea, R., Trkola A., et al. (1994). Cross-neutralizing antibodies against human immunodeficiency virus type 1 isolates induced by the gp41 sequence ELDKWAS. J Virol 68 4031-4034.

Myers, G., Korber, B., Berzofsky, J., Smith, T. F., and Pavlakis, G. N. (1991). Human Retroviruses and AIDS 1991. Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

Myers, G., MacInnes, K., Korber, B. (1992). The emergence of simian/human immunodeficiency viruses. AIDS Res Hum Ret 8 373-86.

Nabel, G, Baltimore, D. (1987). An inducible transcription factor activates expression of human immunodeficiency virus in T cells. Nature 326 711-713.

Nara, P., Dunlop, L., Hatch, W., et al. (1990). Emergence of viruses resistant to neutralization by V3-specific antibodies in experimental human immunodeficiency virus type 1 IIIB infection of chimpanzees. J Virol 64 3779-3791.

Narayan, O., Griffin, D., Chase, J. (1977). Antigenic shift of visna virus in persistantly infecetd sheep. Science 197 376-378.

Niu, M., Stein, D., and Schnittman, S. (1993). Primary human immunodeficiency virus type 1 infection: review of pathogenesis and early treatment intervention in humans and animal retrovirus infections. J Infect Dis 168 1490-1501.

Nowak, M. (1992). What is a quasispecies? TREE 7 118-121.

Nowak, M., May, R., Anderson, R. (1990). The evolutionary dynamics of HIV-1 quasispecies and the development of immunodeficiency disease. AIDS 4 1095-1103.

Nowak, M and May, R. (1993). AIDS Pathogenesis: mathematical models of HIV and SIV infections. AIDS S3-18.

O'Brien, W, Koyangi, Y., Namazie, A., et al. (1990). HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding site. Nature 348 69-73.

Oka, S., Ida, S., Shioda, T., Takebe, Y., et al. (1994). Genetic analysis during a rapid progression to AIDS in an apparently healthy man. AIDS Res Hum Ret 10 271-277

Olshevsky, U., Helseth, E., Furman, C., et al. (1990). Identification of individual human immunodeficiency virus type 1 gp120 amino acids important for Cd4 receptor binding. J Virol 64 5701-5707.

Ou, CY., Kwok, S., Mitchell, S., et al. (1988). DNA amplification for direct detection of HIV-1 in DNA of peripheral blood mononuclear cells. Science 239 295-297.

Ou, C., Ciesielski, C., Myers, G., et al. (1992). Molecular epidemiology of HIV transmission in a dental practice. Science 256 1165-1171.

Page, K., Stearns, S., Littman. (1992). Analysis of mutations in the V3 domain of gp160 that affect fusion and infectivity. J Virol 66 524-533.

Palker, T., Clark, M., Langlois, A., Matthews, T., et al. (1988). Type-specific neutralization of the human immunodeficiency virus with antibodies to env encoded synthetic peptides. PNAS USA 85 1932-1936.

Pang, S., Vinters, H., Akashi, T., O'Brien, W., Chen, I. (1991). HIV-1 Env sequence variation in brain tissue of patients with AIDS-related neurological disease. J Acq Imm Def Synd 4 1082-1092

Pantaleo, G., Graziosi, C., Demarest, J., et al. (1993). HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of infection. Nature 362 355-358.

Papsidero, L., Sheu, M., Ruscetti, F. (1989). Human immunodeficiency virus type 1-neutralising monoclonal antibodies which react with p17 core protein: characterisation and epitope mapping. J Virol 63 267-272.

Pathak, V., and Temin, H. (1990a). Broad spectrum of in vivo mutations, hypermutations, and mutational hotspots in a retrviral shuttle vector after a single replication cycle: substitutions, frameshifts and hypermutations. PNAS USA 87 6019-6023.

Pathak, V., and Temin, H. (1990b). Broad spectrum of in vivo mutations, hypermutations, and mutational hotspots in a retrviral shuttle vector after a single replication cycle: Deletions and deletions with insertions. PNAS 87 6024-6028.

Patterson, S., knight, S. (1987). Susceptibility of human peripheral blood dendritic cells to infection by human immunodeficiency virus. J Gen Virol 68 1177-1181.

Pelchem-Matthews, A., Clapham, P., and Marsh, M. (1995). Role of CD4 endocytosis in human immunodeficiency virus infection. J Virol 69 8164-8168.

Perez, L., O'Donnell, M., Stephens, E. (1992). The transmembrane glycoprotein of human immunodeficiency virus type 1 indeuces syncytium formation in the absence of the receptor binding glycoprotein. J Virol 66 4134-4143.

Piatek, M., Saag, M., Yang, L., et al. (1993). High levels of HIV-1 in plasma during

all stages of infection determined by competitive PCR. Science 259 2505-2513.

Pinter, A., Honnen, W., Tilley, S., Bona, C., et al. (1989). Oligomeric structure of gp41, the transmembrane protein of human immunodeificency virus type 1. J Virol 63 2674-2679.

Pomerantz, R., Feinberg, A., Andino, R., Baltiomore, D. (1991). The long terminal repeat is not a major determinant of the cellular tropism of human immunodeficiency virus type 1 . J Virol 65 1041-1045.

Pope, M., Romani, N., Hirmand, H., et al. (1994). Conjugets of dendtritic cells and memory T lymphocytes from skin facilitate productive infection with HIV-1. Cell 78 389-398.

Posner, M., Hideshima, T., Cannon, T., et al. (1991). An IgG human monoclonal antibody that reacts with HIV-1/gp120, inhibits virus binding to cells, and neutralises infection. J Immunol 146 4325-4332.

Power, C., McArthur, J., Griffin, D., et al. (1994). Demented and nondemented patients with AIDS differ in brain-derived human immunodeficiency virus type 1 envelope sequences. J Virol 68 4643-49.

Preston, B., Poiesz, B., Loeb, L. (1988). Fidelity of HIV-1 reverse transcriptase. Science 242 1168-1171.

Price, R., Brew, B., Sidtis, J., et al. (1988). The brain in AIDS: central nervous system HIV-1 infection and AIDS dimentia complex. Science 239 586-596.

Profy, A., Salinas, P., Eckler, L., Dunlop, N., et al. (1990). Epitopes recogniseed by the neutralising antibodies of an HIV-1-infected individual. J Immunol 144 4641-4647.

Putney, S., Matthews, T., Robey, G., Lynn, D., et al (1986). HTLV-III/LAV-neutralising antibodies to an E. coli-produced fragment of the virus envelope. Science 234 1392-1395.

Ratner, L., Haseltine, W., Patarca, et al. (1985). Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature 313 277-284.

Ratner, L. (1992). Glucosidase inhibitors for treatment of HIV-1 infection. AIDS Res Hum Ret 8 165-73.

Reitz, M., Hall, L., Robert-Guroff, M., et al. (1994). Viral variability and serum antibody response in a laboratory worker infected with HIV type 1 (HTLV IIIB). AIDS Res Hum Retroviruses 10 1143-1155.

Robert-Guroff, M., Brown, M., Gallo., R. (1985). HTLV-III-neutralising antibodies in patients with AIDS and AIDS-related complex. Nature 316 72-74.

Robert-Guroff, M., Reitz, M., Robey, W., and Gallo, R. (1986). In vitro generation of an HTLV-III variant by neutralising antibody. J Immunol 137 3306-3309.

Roberts, J., Bebenek, K., Kunkel, T. (1988). The accuracy of reverse transcriptase from HIV-1. Science 242 1171-1173.

Robey, W., Safai,B., Oroslan, S., Arthur, L., et al. (1985).Characterisation of envelope and core structural gene products of HTLV-III with sera from AIDS patients. Science 228 593-295.

Rusche, J., Javaherian, K., McDanal, C., et al. (1988). Antibodies that inhibit fusion of HIV-1 infected cells bind to 24 amino acid sequence of the viral envelope, gp120. PNAS 85 3198-3204.

Russel, M., Kidd, S. and Kelley, M. (1986). An improved filamentous helper phage for generating single-stranded plasmid DNA. Gene 42 333-338.

Rutherford, G., Lifson, A., Hessol, N., et al. (1990). Course of HIV-1 infection in a cohert of homosexual and bisexual men: an 11 year follow up study. Br Med J 301 1183-1188.

Saag, M., Hahn, B., Gibbons, J., Li, Y., et al. (1988). Extensive variation of human immunodeficiency virus type 1 in vivo. Nature 334 440-444.

Sabin, A. (1992). Imbrobability of effective vaccination against human immunodeficiency virus because of its intracellular transmission and rectal port of entry. PNAS USA 89 8852-8855.

Sabino, E., Pan, LZ., Cheng-Mayer, C., Mayer, A. (1994). Comparison of in vivo plasma and peripheral blood mononuclear cell HIV-1 quasispecies to short-term tissue culture isolates: an analysis of tat and C2-V3 env regions. AIDS 8 901-909.

Saiki, R., Scharf, S., Faloona, F., Mullis, K., Horn, G., Erlich H., Arnheim, N (1985). Enzymatic amplification of β-globin genomic sequences and restriction site analysis for sickle cell anemia. Science 230 1350-1354.

Saiki, R., Gelfand D., Stoffel, S., Scharf, S., Higuchi, R., et al. (1988). Primer-specific enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239 487-91.

Sakai, K., Dewhurst, S., Ma,X., Volsky, D. (1988). Differences in cytopathogenicity and host cell range among infectious molecular clones of human immunodeficiency virus type 1 simultaneously isolated from an individual. J Virol 62 4078-4085.

Salahuddin, S., Markham, P., Wong-Staal, F., et al. (1983). Restricted expression of human T-cell leukemia-lymphoma virus (HTLV) in transformed human umbilical cord blood lymphocytes. Virology 129 51-64.

Salinovich, O., Payne, S., Montelaro, R., et al. (1986). Rapid emergence of novel

antigenic and genetic variants of equine infectious anaemia virus during persistant infection. J Virol 57 71-80.

Salminen,M., Nykanen, A., Brummer-Korvenkontio, H., et al. (1993). Molecular epidemiology of HIV-1 based on phylogenetic analysis of in vivo gag p7/p9 direct sequences. Virology 195 185-194.

Sanchez-Palomino, S., Rojas, J., Martinez, M., et al. (1993). Dilute passage promtes expression of genetic and phenotypic varaints of human immunodeficiency virus type 1 in cell culture. J Virol 67 2938-2943.

Sanger, F., Nicklen, S. and Coulson, R. (1977). DNA sequencing with chain terminating inhibitors. PNAS USA 74 5463-5467.

Sattentau, Q., and Moore, J. (1991). Conformational changes induced in human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. J Exp Med 174 407-415.

Sattentau, Q., Arthos, J., Deen, K., Haana, D., et al. (1989). A structural analysis of the human immunodeficiency virus-binding domain of CD4. Epitope mapping with site-directed mutants and anti-idiotypes. J Exp Med 170 1319-1334.

Sattentau, Q., Zolla-Pazner, S., Poignard, P. (1995). Epitope exposure on functional, oligomeric gp41 molecules. Virology 206 713-717.

Scarlatti, G., Leitner, T., Halapi, E., Wahlberg, J. et al. (1993). Comparison of variable region 3 sequences of human immunodeficiency virus type 1 from infected children with the RNA and DNA sequences of the virus populations of their mothers. PNAS USA 90 1721-1725.

Schnittman, S., Psallidopoulos, M., Lane, H., Thompson, L. et al. (1989). The reservoir for HIV-1 in human peripheral blood is a T cell that maintains expression of CD4. Science 245 305-308.

Schnittman, S., Lane, H., Greenhouse, J., et al. (1990). Preferential infection of CD4+ memory T cells by human immunodeficiency virus type 1: evidence for a role in the selective T-cell functional defects observed in infected individuals. PNAS USA 87 6058-6062.

Schuitemaker, H., Kootsra, N., de Goede, R., et al. (1991). Monocytotropic human immunodeficiency virus type 1(HIV-1) variants detectable in all stages of infection lack T-cell line tropism and syncytium-inducing ability in primary T-cell culture. J Virol 65 356-363.

Schuitemaker, H., Koot, M., Koostra, N., et al. (1992). Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression to disease is associated with a shift from monocytotropic to T-cell-tropic virus populations. J Virol 66 1354-1360.

Schuitemaker, H., Meyaard, L., Kootstra, N., et al. (1993). Lack of T-cell dysfunction and programmed cell death in human immunodeficiency virus type 1 infected chimpanzeescorrelates with the absence of monocytotropic variants. J Infect Dis 168 1140-1147.

Schulz, T., Jameson, B., Lopalco, L., et al. (1992). Conserved structural features in the interaction between retroviral surface and transmembrane glycoproteins? AIDS Res Hum Retroviruses 8 1571-1580.

Schwartz, S., Felber, B., Fenyo, EM., Pavlakis, G. (1989). Rapidly and slowly replicating human immunodeficiency virus type 1 isolates can be distinguished according to target-cell tropism in T-cell and monocyte cell lines. PNAS USA 86 7200-7203.

Schwartz, S., Felber, B., Fenyo, EM., Pavlakis G. (1990a). Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. J Virol 64 2519-2529.

Schwartz, S., Felber, B., Fenyo, EM., Pavlakis, G. (1990b). Env and Vpu protein of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. J Virol 64 5448-5456.

Schwartz, O., Marechal, V., Danos, O., and Heard, J. (1995). Human immunodeficiency virus type 1 Nef increases the efficiency of reverse transcription in the infected cell. J Virol 69 4053-4059.

Shaw, G., Hahn, B., Arya, S., et al. (1984). Molecular characterisation of human T-cell leukemia (lymphotropic) virus type III in the acquired immune deficiency syndrome. Science 226 1165-1171.

Shimizu, N., Shimizu, N, Takeuchi, Y., Hoshino., N. (1994). Isolation and characterization of human immunodficiency virus type 1 variants infectious to brain-derived cells: detection of common point mutations in the V3 region of the env gene of the variants. J Virol 68 6130-6135.

Shioda, T., Levy, J., Cheng-Mayer, C. (1991). Macrophage and T cell-line tropisms of HIV-1 are determined by specific regions of the envelope gp120 gene. Nature 349 167-169.

Short, J., Fernandez, J., Sorge, J., and Huse, W. (1988). Lambda ZAP: a bacteriophage lambda expression vector with *in vivo* excision properties. Nucleic Acid Research 16 7583-7600.

Shpaer, E., Mullins, J. (1993). Rates of amino acid change in the envelope protein correlate with pathogenicity of primate lentiviruses. J Mol Evol 37 57-65.

Simmonds, P., Balfe, B., Peutherer, et al. (1990a). Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. J Virol 64 864-872.

Simmonds, P., Balfe, P., Ludlam, et al. (1990b). Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. J Virol 64 5840-5850.

Simmonds, P., Zhang, L.Q, McOmish, et al. (1991) Discontinuous sequence change of human immunodeficiency virus (HIV) type 1 env sequences in plasma viral and lymphocyte-associated proviral populations in vivo: implications for models of HIV pathogenesis. J Virol 65 6266-6276.

Skinner, M., Langlois, A., McDanal, C., et al (1988). Neutralizing antibodies to an immunodominant envelope sequence do not prevent gp120 binding to CD4. J Virol 4195-4200.

Smith, S., Shatsky, M., Cohen, P., et al. (1984). Monoclonal antibody and enzyme profiles of human malignant T-lymphoid cells and derived cell lines. Cancer Research 44 5657-5660.

Sodrosky, J., Goh, W., Rosen, C., Campbell, K., Haseltine, W. (1986). Role of the HTLV/LAV envelope in syncytiumformation and cytopathicity. Nature 322 470-474.

Southgate, C., Zaapp, M., Green, M. (1990). Activation of transcription by Tat protein tethered to nascent RNA through another protein. Nature 345 640-642.

Spina, C., Kwoh, T., Chowers, M., et al. (1994). The importance of nef in the induction of human immunodeficiency virus type 1 replication from primary quiescent CD4 lymphocytes. J Exp Med 179 115-123.

Spira A., Ho, D. (1995). Effect of different donor cells on human immunodeficiency virus type 1 replication and selection in vitro J Virol 69 422-429.

Stamatatos, L., and Cheng-Mayer, C. (1993). Evidence that the structural conformation of the envelope gp120 affects human immunodeficiency type 1 infectivity, host-range, and syncytium-forming ability. J Virol 67 5635-5639.

Stamatatos, L., Werner, A., Cheng-Mayer, C. (1994). Differential regulation of cellular tropism and sensitivity to soluble CD4 neutralization by the envelope gp120 of human immunodeficiency virus type 1. J Virol 68 4973-4979.

Starcich, B., Hahn, B., Shaw, G., et al. (1986). Identification and characterization of conserved and variable regions in the envelope gene of HLTV-III/LAV, the retrovirus of AIDS. Cell 45 637-648.

Steimer, K., Scandella, C., Stiles, P, Haigwood, N. (1991). Neutralization of divergent HIV-1 isolates by conformation-dependent human antibodies to gp120. Science 254 105-108.

Stein, B., Gowda, S., Lifson, J., et al. (1987). pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. Cell 49 659-668.

Steinhauer, D., and Holland, J. (1987). Rapid evolution of RNA viruses. Ann Rev Microbiol 41 409-33.

Stevenson, M., Stanwick, T., Dempsey, M., and Lamonica, C. (1990a). HIV-1 replication is controlled at the level of T cell activation and proviral integration. EMBO Journal 9 1551-1560.

Stevenson, M., Haggerty, S., Lamonica, C., Meier, C., et al. (1990b). Integration is not necessary for expression of human immunodeficiency virus type 1 protein products. J Virol 64 2421-2425.

Stott, E., Kitchin, P., Page, M., et al. (1991). Anti-cell antibodies in macaques. Nature 353 393.

Strebel, K., Klimkait, T., Martin, M. (1988). A novel gene of HIV-1, vpu, and its 16-kilodalton product. Science 241 1221-1223.

Studier, W. and Moffatt, B. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol 189 113-130.

Sullivan, N., Thali, M., Furman, C., Ho, D., Sodroski, J. (1993). Effect of amino acid changes in the V1/V2 region of the human immunodeficiency virus type 1 gp120 glycoprotein on subunit association, syncytium-formation and recognition by neutralizing antibody. J Virol 67 3674-3679.

Sullivan, N., Sun, Y., Li, J, Hoffman, W., Sodroski, J. (1995). Replicative function and neutralization sensitivity of envelope glycoproteins from primary and T cell-line passaged human immunodeficiency virus type 1 isolates. J Virol 69 4413-4422.

Sundstrom, C., and Nilsson, K. (1976). Establishment and characterisation of a human histiocytic lymphoma cell line (U-937). J Cancer 17 565-577.

Takeda, A., Tuazon, C., Ennis, F. (1988). Antibody-enhanced infection by HIV-1 via Fc recptor mediated entry. Science 242 580-583.

Takeuchi, Y., Akutsu, M., Murayama, K., Shimizu, N., Hoshino, H. (1991). Host range of human immunodeficiency virus type 1 : modification of cell tropism by a single point mutation at the neutralization epitope in the env gene. J Virol 65 1710-1718.

Temin, H. (1993). A proposal for a new approach to a preventative vaccine against human immunodeficiency virus type 1. PNAS USA 90 4419-4420.

Tersmette, M. and Schuitemaker, H. (1993). Virulent HIV strains ? AIDS 7 1123-1125.

Tersmette, M., De Goede, R., Al, B., et al. (1988). Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency (AIDS) and AIDS-

related complex. J Virol 62 2026-2032.

Tersmette, M., Lange, J., de Goede, R., et al. (1989a). Association between biological properties of human immunodeficiency virus variants and risk for AIDS and AIDS mortality. Lancet i 983-985.

Tersmette, M., Gruters, R, de Wolf, F., et al. (1989b). Evidence for a role of virulent human immunodeficiency virus (HIV) variants in the pathogenesis of acquired immunodeficiency syndrome: studies on sequential isolates. J Virol 63 2118-2125.

Terwilliger, E., Cohen, E., Lu, Y., Sodroski, J., Haseltine, W. (1989). Functional role of human immunodeficiency virus type 1 vpu. PNAS USA 86 5163-5267.

Thali, M., Moore, J., Furman, C., et al. (1993). Characterisation of human immunodeficiency virus type 1 gp120 neutralising epitopes exposed upon gp120-CD4 binding. J Virol 67 3908-3988.

Thali, M., Charles, M., Furman, C. et al. (1994). Resistance to neutralization by broadly reactive antibodies to the human immunodeficiency virus type 1 gp120 glycoprotein conferred by a gp41 amino acid change. J Virol 68 674-680.

Thormar, H., Barshatzky, M., Arnessen, K., Kozlowski, P. (1983). The emergence of antigenic variants is a rare event in long-term visna virus infection in vivo. J Gen Virol 64 1427-1432.

Tindall, K., and Kunkel, T. (1988). Fidelity of DNA synthesis by Thermophilus aquaticus DNA polymerase. Biochemistry 28 6008-6013.

Tracy, S.(1981). Improved rapid methodology for the isolation of nucleic acids from agarose gels. Preparative Biochemistry 11 251-268.

Tremblay, M, Wainberg, M. (1990). Neutralisation of multiple isolates from a sinlge subject by autologous sequential sera. J Inf Dis 162 735-737.

Valentin, A., Albert, J., Fenyo, E.M., Asjo, B. (1994). Dual tropism for macrophages and lymphocytes is a common feature of primary human immunodeficiency type 1 and 2 isolates. J Virol 68 6684-6689.

Varmus, H. (1988). Retroviruses. Science 240 1427-1435.

Vartanian, J., Meyerhans, A., Asjo, B., Wain-Hobson, S. (1991). Selection, recombination and G -> A hypermutation of human immunodeficiency virus type 1 genomes. J Virol 65 1779-1788.

Venet, M., and Walker, B. (1993). Cytotoxic T-cell epitopes in HIV/SIV infection. AIDS S117-126

von Schwedler, U., Song, J., Aiken, C., and Trono, D. (1993). Vif is crucial for HIV-1 proviral DNA synthesis in infected cells. J Virol 67 4945-4955.

Wahlberg, J., Albert, J., Lundeberg, J., et al. (1991). Analysis of the V3 loop in neutralization-resistant human immunodeficiency virus type 1 variants by direct solid phase DNA sequencing. AIDS Res Hum Retroviruses 7 983-990.

Wain-Hobson, S., Vartanian, J-P., Henry, M., et al. (1991). LAV revisited: origins of the early HIV-1 isolates from Institut Pasteur. Science 252 961-965.

Webster, R., Laver, W., Air, G., Schild, G. (1982). Molecular mechanisims of variation in influenza viruses. Nature 296 115-121.

Weinberg, J., Matthews, T., Cullen, B., and Malim, M. (1991). Productive human immonodeficiency virus type 1 (HIV-1) infection of nonproliferating human monocytes. J Exp Med 174 1477-1482.

Weiss. R., Clapham, P., Cheinsong-Popovic, R., et al. (1985a). Neutralisation of human T-lymphotropic virus type III by sera of AIDS and AIDS-risk patients. Nature 316 69-72.

Weiss, R., Teich, N., Varmus, H., Coffin, J. (1985b). *RNA Tumor Viruses* 2nd edition vols. 1 and 2. Cold Spring Harbor Laboratory, Cold Spring Harbour, New York.

Weiss, R., Clapham, P., Weber, J., Dalgliesh, A., et al. (1986). Variable and conserved neutralisation antigens of human immunodeficiecny virus. Nature 324 572-575.

Westervelt, P., Gendelman, H., Ratner, L. (1991). Identification of a determinant within the human immunodeficiency virus type 1 surface envelope glycoprotein critical for productive infection of primary monocytes. PNAS USA 88 3097-3101.

Westervelt, P., Trowbridge, D., Epstein, L., et al. (1992a). Macrophage determinants of human immunodeficiency virus type 1 in vivo. J Virol 66 2577-2582.

Westervelt, P., Henkel, T., Trowbridge, D., et al. (1992b). Dual regulation of silent and productive infection in monocytes by distinct human immunodeficiency virus type 1 determinants. J Virol 66 3925-3931.

White, J.M. (1990). Viral and cellular fusion proteins. Ann Rev Biochem 56 365-394

White, F. and Fenner, F. (1994). *Medical Virology 4th Edition.* The Academic Press ltd, 24-28 Oval Road, London NW1 7DX.

Wiley, C., Schrier, R., Nelson, J., et al (1986). Cellular localization of HIV infection within the brain of AIDS patients. PNAS USA 83 7089-7093.

Willey, R., Rutledge, R., Dias, S., et al. (1986). Identification of divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus. PNAS USA 83 5038-5042.

Willey, R., Bonifacino, J., Potts, B., et al. (1988). Biosynthesis, cleavage, and degradation of the human immunodeficiency virus 1 envelope glycoprotein gp160. PNAS USA 85 9580-9584.

Willey, R., Maldarelli, F., Martin, M., Strebel, K. (1992a). Human immunodeficiency virus type 1 Vpu protein regulates the formation of intracellular gp160-CD4 complexes. J Virol 66 226-234.

Willey, R., Maldarelli, F., Martin, M., Strebel, K. (1992b). Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4. J Virol 66 7193-7200.

Willey, R, Theodore, T., Martin, M. (1994). Amino acid substitutions in human immunodeficiency virus type 1 gp120 V3 loop that change viral tropism also alter physical and functional properties of the viral envelope. J Virol 68 4409-4419.

Wilson, C., Reitz, M., Aldrich, K., Klasse, P. et al. (1990). The site of an immune-selected point mutation in the transmembrane protein of human immunodeficiency virus type 1 does not constitute the neutralization epitope. J Virol 64 3240-3248.

Wilson, W., Braddock, M., Adams, S., et al. (1988). HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. Cell 55 1159-1169.

Winship, P. R. (1989). An improvised method for directly sequencing PCR amplified material using dimethylm sulfoxide. Nucleic Acid Research 17 1266.

Wolfs, T., De Jong, JJ., Van den Berg, H., et al. (1990). Evolution of sequences encoding the principal neutralizing epitope of human immunodefiency virus type 1 is host dependent, rapid, and continuous. PNAS USA 87 9938-9942.

Wolfs, T., Zwart, G., Bakker, M., Cuiken, C., et al. (1991). Naturally occurring mutations within the HIV-1 genomic RNA lead to antigenic variation dependent upon single amino acid substitution. Virology 185 195-205.

Wolfs, T., Zwart, G., Bakker, M., Goudsmit, J. (1992). HIV-1 genome RNA diversification following sexual and parenteral virus transmission. Virology 189 103-110.

Wolinsky, S., Wike, C., Korber. B., et al. (1992). Selective transmission of human immunodeficiency virus type 1 variants from mothers to infants. Science 255 1134-1137.

World Health Organization. (1993). Statistics from the World Health Organization and the Centres for Disease Control and Prevention. AIDS 7 1287-1291.

World Health Organisation. (1994). Feasibility of developing live attenuated HIV vaccines: conclusions and recommendations. AIDS Res Hun Ret 10 221-222.

Wrin, T., Loh, T., Vennari, J., Schuitrmaker, H., Nunberg, J. (1995). Adaptation to

growth in H9 cell line renders a primary isolate of HIV-1 sensitive to neutralization by vaccine sera. J Virol 69 39-48.

Wyatt, R., Thali, M., Tilley, S., et al. (1992). Relationship of the human immunodeficiency virus type 1 gp120 third variable loop to elements of the CD4 binding site. J Virol 66 6997-7004.

Yamada, O., Matsumoto, T., Sasoaka, R., Kurimura, T. (1990). Variations in growth capacity of HIV in peripheral blood mononuclear cell preparations from different individuals. AIDS 4 35-40.

Yamashita, A., Yamamoto, N., Matsuda, J., Koyanagi, Y. (1994). Cell type-specific heterogeneity of the HIV-1 V3 loop in infected individuals: selection of virus in macrophages and plasma. Virology 204 170-179.

York-Higgins, D., Cheng-Mayer, C., Levey, J., Dina, D. (1990). Human immunodeficiency virus type 1 cellular host range, replication, and cytopathicity are linked to the envelope region of the viral genome. J Virol 64 4016-4020.

Yoshiyama, H., Mo, H., Moore, J., and Ho, D. (1994). Characterisation of mutants of human immunodeficiency virus type 1 that have escaped neutralisation by a monoclonal antibody to the gp120 V2 loop. J Virol 68 974-978.

Yourno, J., Fisher, A., Looney, D., Gallo, R., Wong-Staal, F. (1989). A recombinant clone of HIV-1 preferentially transmitted in normal peripheral blood mononuclear cells. AIDS Res Hum Retroviruses 5 565-575.

Zack, J,. Arrigo, S., Weitsman, A., et al. (1990). HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure. Cell 61 213-222.

Zhang, LQ. (1993). *The Pattern and Process of Human Immunodeficiency Virus Sequence Evolution In Vivo.* PhD thesis. The University of Edinburgh.

Zhang, LQ., Simmonds, P., Ludlam, C., Leigh Brown, A. (1991). Detection, quantification and sequencing of HIV-1 from the plasma of seropositive individuals and from factor VIII concentrates. AIDS 5 675-681.

Zhang, L. Q., MacKenzie, P., Clelland, A., et al. (1993). Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. J Virol 67 3345-3356.

Zhang, L-Q., Diaz, R., Ho, D. et al. (1996). Host-specific driving force in HIV-1 evolution in vivo. Submitted.

Zhu, T., Mo, H., Wang, N., et al. (1993). Genotypic and phenotypic characterization of HIV-1 in patients with primary infection. Science 261 1179-1181.

Zwart, G., Langedijk, H., van der Hoek, L., et al. (1991). Immunodominance and

antigenic variation of the principal neutralisation domain of HIV-1. Virology 181 481-489.

Zwart, G., Wolfs, T., Valk, M., Van Der Hoek, L., et al. (1992). Characterisation of the specificity of the human antibody responses to the V3 neutralisation domain of HIV-1. AIDS Res Hum Ret 8 1897-1908.

Zwart, G., Van Der Hoek, L., Valk, M., et al. (1994). Antibody responses to HIV-1 Envelope and Gag epitopes in HIV-1 seroconverters with rapid versus slow disease progression. Virology 201 285-293.